

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Статистическое моделирование

Охотников Григорий Иванович

ПРИМЕНЕНИЕ МЕТОДА АНАЛИЗА СИНГУЛЯРНОГО СПЕКТРА

Выпускная квалификационная работа

Научный руководитель:

к. ф.-м. н., доцент Н. Э. Голяндина

Рецензент:

к. ф.-м. н., лектор А. Н. Пепельшев

Санкт-Петербург

2017

Saint Petersburg State University
Applied Mathematics and Computer Science
Statistical Modelling

Okhotnikov Grigorii Ivanovich

APPLICATION OF SINGULAR SPECTRUM ANALYSIS

Graduation Project

Scientific Supervisor:

Associate Professor N. E. Golyandina

Reviewer:

Lecturer A. N. Pepelyshev

Saint Petersburg

2017

Оглавление

Введение	4
Глава 1. Метод «Гусеница»-SSA	6
1.1. Разложение	6
1.2. Восстановление	7
1.3. Прогнозирование	8
Глава 2. Анализ данных	9
2.1. Исходные данные	9
2.2. Предварительный анализ временных рядов	10
Глава 3. Прогнозирование рядов ПВЗ	19
3.1. Подбор параметров разложения и восстановления	19
3.1.1. Выбор параметров L и r	23
3.1.2. Выбор длины базового отрезка	25
3.1.3. Выбор длины отрезка кросс-валидации	26
3.2. Сравнение результатов прогнозирования	27
3.3. Описание интерфейса веб-приложения	28
3.3.1. Страница «Forecast for Today»	31
3.3.2. Страница «Generate Forecast»	31
3.3.3. Страница «Compare Forecasts»	32
3.3.4. Страница «About»	32
Заключение	41
Список литературы	42
Приложение А. Таблицы значений параметров L^* и r^*	44
Приложение Б. Графики прогнозов для различных начальных дат	51

Введение

Изучение вращения Земли представляет большой интерес для современных геологов и астрономов. Определение положения Земли в космосе относительно удалённых радиоизлучающих неподвижных точек позволяет с большой точностью определять орбиты спутников и координаты пользователей GPS.

Как правило, наибольший интерес представляют пять параметров вращения Земли (ПВЗ; EOP, Earth orientation parameters): координаты полюса в земной системе координат x и y ; долгота дня LOD (length of day) и отклонение небесного полюса dX и dY . Полюс (x, y) находится в точке поверхности Земли, где проходит мгновенная ось вращения планеты. Он подвержен годичному и чандлеровскому колебаниям, приливному действию Луны, воздействию ветров и океанских течений и прочим факторам. Долгота дня LOD отражает разницу между фактическим временем, за которое планета совершила полный оборот, и 86400 секундами. Долгота дня является первой производной ряда $UT1 - UTC$ отклонений среднего времени по Гринвичу $UT1$ от всемирного координированного времени UTC (с учётом всех секунд координации). Отклонение (dX, dY) является поправкой модели прецессии-нутации IAU 2000, определяющей положение так называемого небесного полюса, т. е. являются отличием наблюдаемого положения полюса от его модельного значения, поэтому оно похоже на шум. Привязка значений рядов по дням осуществляется по модифицированной юлианской дате (MJD, Modified Julian Date). Наблюдением ПВЗ занимается Международная служба вращения Земли (МСВЗ; IERS, International Earth Rotation and Reference Systems Service). Она же фиксирует стандарты земной и небесной систем координат, описанные в специальном документе IERS Conventions (2010). Кроме точных значений параметров вращения Земли важное практическое применение имеют спрогнозированные значения некоторых из параметров. Прогнозированием занимается отделение службы вращения Земли, базирующееся в Военно-морской обсерватории США (USNO, United States Naval Observatory).

В результате проведения кампании по предсказанию параметров вращения Земли EOP PCC (Earth orientation parameters prediction comparison campaign) [1], проходившей с 2005 по 2008 годы при координировании Венского технического университета, был сделан вывод о том, что прогноз, получаемый в виде комбинации прогнозов, выполненных различными методами, оказывается лучше, чем каждый из прогнозов в отдельности.

Впоследствии подобный эксперимент был повторён самой USNO [2]. Результаты прогнозов, предоставленных участниками, хранятся в архиве и доступны на сайте проекта, однако с февраля 2015 года базы данных не обновляются.

Попытки применения метода SSA к задаче прогнозирования параметров вращения Земли предпринимались и ранее [3, 4, 5]. Однако перед авторами исследований стояли технические ограничения, которые были преодолены с выходом пакета RSSA [6, 7], а опубликованные результаты не удаётся воспроизвести и использовать для ежедневного прогнозирования упомянутых временных рядов.

Целями данной работы является прогнозирование каждого из пяти рядов параметров вращения Земли, создание алгоритма прогнозирования этих рядов и ежедневная публикация на странице в Интернете результатов его работы, а также сравнение результатов прогноза с результатами, полученными другими методами.

Метод «Гусеница»-SSA

Ниже приводятся основные этапы алгоритма «Гусеница»-SSA по статье [8].

Рассмотрим вещественнозначный временной ряд $\mathbb{X}_N = (x_1, \dots, x_N)$ длины N . Пусть L ($1 < L < N$) — некоторое целое число (длина окна), $K = N - L + 1$. Алгоритм метода «Гусеница»-SSA состоит из двух этапов: разложения и восстановления.

1.1. Разложение

На этапе разложения образуются $K = N - L + 1$ векторов вложения

$$X_i = (x_i, \dots, x_{i+L-1})^T, \quad i = 1, \dots, K, \quad (1.1)$$

имеющих размерность L . Траекторной матрицей ряда \mathbb{X}_N называется матрица, составленная из векторов вложения в качестве столбцов:

$$\mathbf{X} = [X_1 : \dots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}. \quad (1.2)$$

По построению матрица \mathbf{X} имеет одинаковые элементы на «диагоналях» $i + j = \text{const}$. Такие матрицы называются ганкелевыми. Существует взаимно-однозначное соответствие между ганкелевыми матрицами размерности $L \times K$ и рядами длины $N = L + K - 1$.

Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Обозначим $\lambda_1, \dots, \lambda_L$ собственные числа матрицы \mathbf{S} , взятые в неубывающем порядке ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) и U_1, \dots, U_L — ортонормированную систему собственных векторов матрицы \mathbf{S} , соответствующих собственным числам.

Пусть $d = \max\{i : \lambda_i > 0\}$. Если обозначить $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, $i = 1, \dots, d$, то сингулярное разложение матрицы \mathbf{X} может быть записано как

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_L, \quad (1.3)$$

где $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$. Каждая из таких матриц \mathbf{X}_i имеет ранг 1. Они называются элементарными матрицами. Набор $(\sqrt{\lambda_i}, U_i, V_i)$ называется i -ой собственной тройкой сингулярного разложения (1.3).

1.2. Восстановление

На основе разложения (1.3) выполняется процедура группировки, которая делит всё множество индексов $\{1, \dots, d\}$, $d = \max\{i : \lambda_i > 0\}$ на m непересекающихся подмножеств I_1, \dots, I_m .

Пусть $I = \{i_1, \dots, i_p\}$. Тогда результирующей матрицей \mathbf{X}_I , соответствующей группе I , называется матрица

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}. \quad (1.4)$$

Такие матрицы вычисляются для $I = I_1, \dots, I_m$, и тогда разложение (1.3) может быть записано как

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (1.5)$$

Процедура выбора множеств I_1, \dots, I_m называется группировкой собственных троек.

На последнем этапе каждая матрица разложения (1.5) преобразуется в ряд длины N .

Пусть \mathbf{X}_{I_k} — результирующая матрица из (1.5). Тогда ее преобразование соответствует усреднению элементов матрицы вдоль «диагоналей» $i + j = \text{const}$.

Пусть \mathbf{Y} — некоторая матрица размерности $L \times K$ с элементами y_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$. Для определённости положим $L \leq K$. Процедура диагонального усреднения («ганкелизации») матрицы преобразует матрицу \mathbf{Y} в ряд (y_1, \dots, y_N) согласно следующему правилу:

$$\tilde{y}_s = \sum_{(l,k) \in A_s} y_{lk} / |A_s|, \quad (1.6)$$

где множества $A_s = \{(l, k) : l + k = s + 1, 1 \leq l \leq L, 1 \leq k \leq K\}$, $s = 1, \dots, L + K - 1$, соответствуют элементам, находящимся на побочных диагоналях, а $|A_s|$ обозначает количество элементов множеств A_s . Иными словами, процесс соответствует усреднению элементов матрицы, находящихся на побочных диагоналях, и последующему составлению этих средних значений во временной ряд.

После применения диагонального усреднения к результирующей матрице \mathbf{X}_{I_k} получается восстановленный ряд $\tilde{\mathbf{X}}^{(k)} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_N^{(k)})$. Таким образом, исходный ряд

(x_1, \dots, x_N) представим в виде суммы m восстановленных рядов:

$$x_n = \sum_{k=1}^m \tilde{x}_n^{(k)}, \quad n = 1, \dots, N. \quad (1.7)$$

Приведённый выше алгоритм метода реализован на языке **R** в пакете **RSSA** [7].

1.3. Прогнозирование

Пусть I — это некоторая группа собственных троек, $U_i \in \mathbb{R}^L$, $i \in I$ — соответствующие собственные векторы, \underline{U}_i — их первые $L-1$ координат и π_i — последняя координата U_i , $\nu^2 = \sum_i \pi_i^2$. Определим $R = (a_{L-1}, \dots, a_1)^T$ как

$$R = \frac{1}{1 - \nu^2} \sum_{i \in I} \pi_i \underline{U}_i. \quad (1.8)$$

Алгоритм рекуррентного прогнозирования состоит из следующих действий:

1. ряд $\mathbb{Y}_{N+M} = (y_1, \dots, y_{N+M})$ определяется как

$$y_i = \begin{cases} \tilde{x}_i & \text{for } i = 1, \dots, N, \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = N+1, \dots, N+M, \end{cases} \quad (1.9)$$

2. числа y_{N+1}, \dots, y_{N+M} являются результатом прогнозирования M точек ряда.

Таким образом, рекуррентный прогноз выполняется непосредственным применением ЛРФ с коэффициентами $\{a_j, j = 1, \dots, L-1\}$.

Возможен иной подход. Введём оператор $\mathcal{P}_{\text{Rec}} : \mathbb{R}^L \mapsto \mathbb{R}^L$ по следующей формуле:

$$\mathcal{P}_{\text{Rec}} Y = \begin{pmatrix} \bar{Y} \\ R^T \bar{Y} \end{pmatrix}. \quad (1.10)$$

Обозначим

$$Z_i = \begin{cases} \tilde{X}_i & \text{for } i = 1, \dots, K, \\ \mathcal{P}_{\text{Rec}} Z_{i-1} & \text{for } i = K+1, \dots, K+M. \end{cases} \quad (1.11)$$

Матрица $\mathbf{Z} = [Z_1 : \dots : Z_{K+M}]$ является траекторной для ряда \mathbb{Y}_{N+M} , и тогда выражение (1.11) можно рассматривать как векторную форму (1.9).

Для получения прогноза, близкого к точным значениям ряда, важно правильным образом образовать группу I , отвечающую за сигнал, присутствующий в ряде. Следует принять во внимание тот факт, что результаты SVD-разложения траекторной матрицы ряда будут, вообще говоря, отличаться при добавлении новых значений к ряду.

Анализ данных

2.1. Исходные данные

Исходными данными для данной работы являются временные ряды параметров вращения Земли, опубликованные МСВЗ в бюллетене С04 [9]. Описание каждого из рядов и процедур их обработки приводится в [10]. Измерения каждого из параметров приводятся ежедневно, с привязкой к модифицированной юлианской дате (MJD). Модифицированная юлианская дата используется Международным астрономическим союзом в качестве стандарта нумерации дней. Например, 1 января 2017 года соответствует $MJD = 57754$.

Из двух рядов UT1–UTC и LOD, присутствующих в С04, будем заниматься анализом ряда LOD, поскольку он приводится в непрерывной шкале без учёта секунд поправки (секунды поправки вводятся Международной службой вращения Земли для уменьшения расхождения между всемирным координированным временем UTC и всемирным временем UT1).

Упомянутый файл бюллетеня С04 содержит ежедневные значения параметров вращения Земли с 1962 года до текущего дня, за исключением последних 30 значений. Эти значения получаются после обработки всех поступивших в МСВЗ данных из различных источников и считаются максимально точными значениями параметров вращения Земли. Бюллетень обновляется ежедневно.

Помимо этого МСВЗ публикует так называемые rapid-данные: оценки ПВЗ, полученные из источников с минимальной задержкой, например, с использованием систем GPS. Это позволяет получить значения ПВЗ вплоть до текущего дня, однако они могут иметь несколько большую погрешность, чем бюллетень С04. Эти файлы имеют суффикс «daily», обновляются ежедневно и содержат оценки ПВЗ за последние 90 дней. Общедоступного архива с файлами daily не ведётся, публикуемые файлы заменяют существующие.

Таким образом, чтобы получить прогноз значений ПВЗ, начиная со следующего дня, следует взять данные из бюллетеня С04, дополнить их с конца данными за последние 30 дней из daily-файла [11] и подать их на вход алгоритму прогнозирования.

Rapid-данные рядов x, y, LOD имеют достаточно высокую точность и не сильно отличаются от значений, публикуемых позднее в C04, поэтому применять какое-либо сглаживание на стыке этих рядов не требуется. Поскольку не существует архива daily-файлов, воспроизвести ситуацию для какого-либо момента в прошлом, аналогичную текущему дню, в которой запаздывание C04 необходимо компенсировать данными из «daily»-файлов, не представляется возможным. В дальнейшем для сравнения результатов прогнозирования для дат в прошлом будем использовать данные из бюллетеня C04, однако следует иметь в виду, что для получения реального прогноза для текущего дня следует использовать данные из двух источников.

2.2. Предварительный анализ временных рядов

Пять параметров ПВЗ условно можно разбить на три группы: координаты полюса x, y , долгота дня LOD и поправки координат небесного полюса dX, dY . Перед прогнозированием рядов следует выполнить их анализ с целью определить группировки, соответствующие сигналу каждого из рядов, а также для того, чтобы учесть их специфику. Выполним SSA-разложение каждого из рядов со значением параметра длины окна L , принятым по умолчанию, т. е. равным половине длины ряда.

Ряды координат полюса x и y приводятся с 1 января 1962 года (рис. 2.1–2.2).

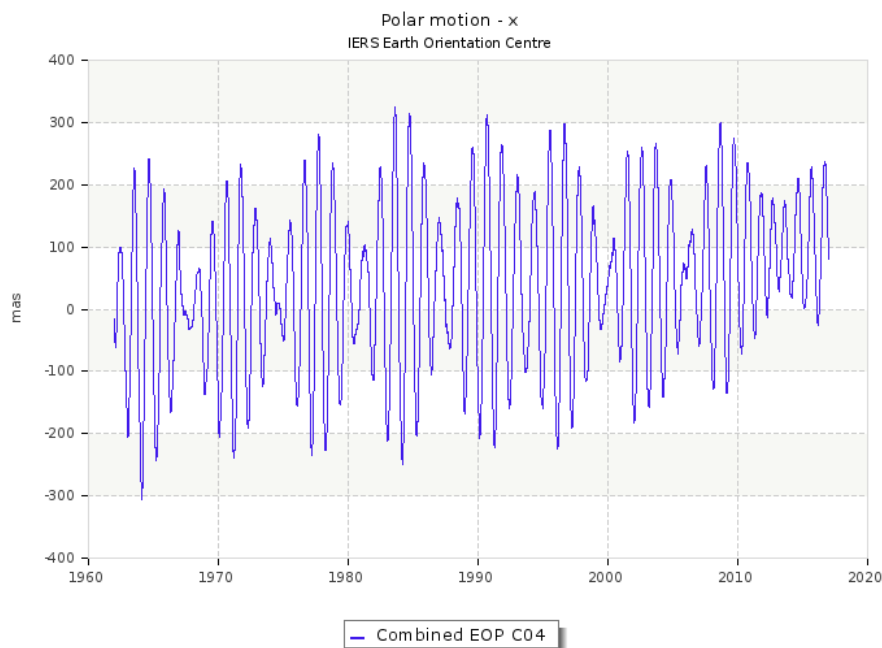


Рис. 2.1. График временного ряда координаты полюса x

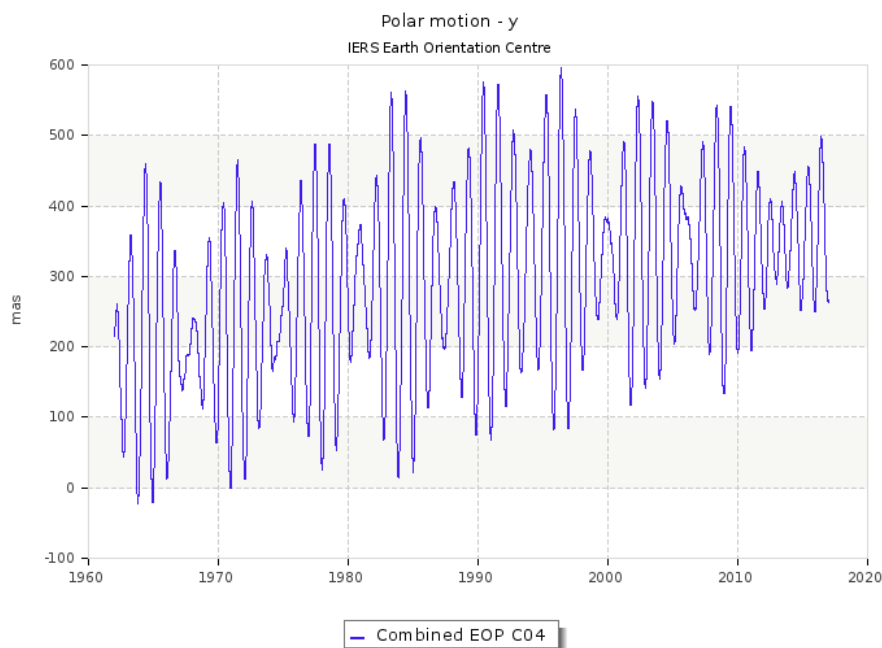


Рис. 2.2. График временного ряда координаты полюса y

Информацию о разнообразии и вкладе частот, присутствующих в ряде, удобно изображать на графике спектрограмм [12].

Спектрограммы рядов x и y (рис. 2.3–2.4) говорят о наличии двух известных составляющих колебания Земли: чандлеровской (приблизительно 432 дня) и годовой (365 дней) [4, 13]. Однако, несмотря на похожесть рядов x и y , их разложения не идентичны. В дальнейшем анализ и прогнозирование этих рядов выполняется независимо друг от друга. Выполним восстановление этих рядов по 21 и 22 первым компонентам соответственно, используя разложения со значением длины окна L , близким к половине длины рядов. Остатки рядов приведены на рис. 2.5–2.6, их спектрограммы — на рис. 2.7–2.8. Из графиков видно, что в силу отсутствия сильной разделимости неизвлечёнными остаются некоторые гармоники.

Ряд LOD приводится также с 1 января 1962 года и является непрерывным, в отличие от ряда $UT1 - UTC$ (рис. 2.9). В нём также присутствуют чандлеровское и годовое колебания. Однако помимо них в ряде содержится большое множество других, более мелких, колебаний различного периода (рис. 2.10). Для их выделения удобно пользоваться парными графиками собственных векторов (рис. 2.11), полученных в результате разложения ряда с длиной окна L , равной половине длины ряда. Существуют такие пары собственных векторов, для которых парные графики представляют собой ровные окружности, что, как правило, свидетельствует о наличии сложного сигнала.

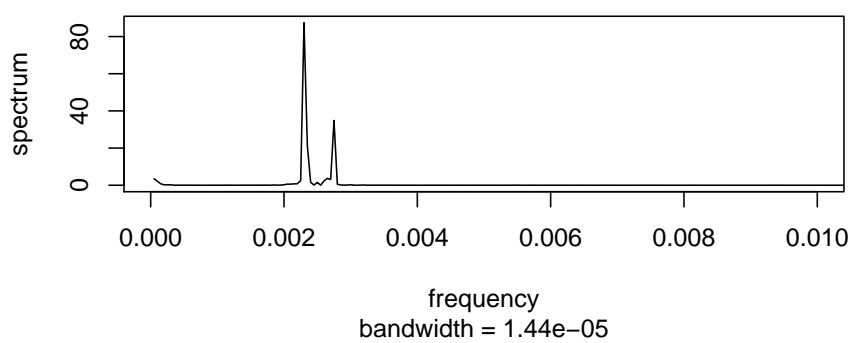


Рис. 2.3. Фрагмент спектрограммы ряда координаты полюса x

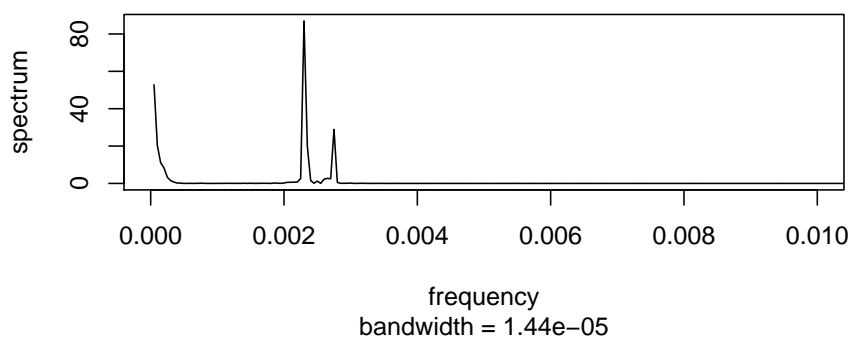


Рис. 2.4. Фрагмент спектрограммы ряда координаты полюса y

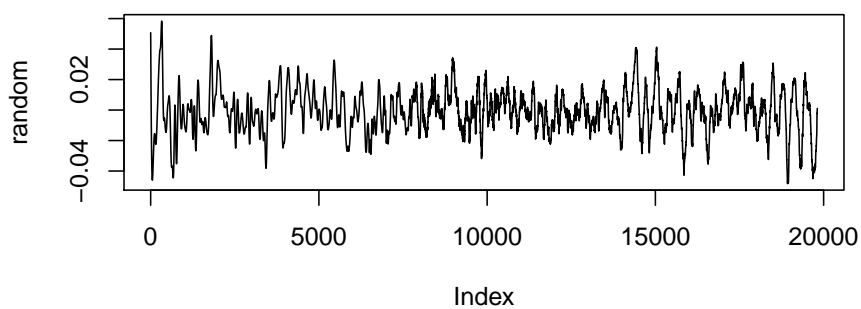


Рис. 2.5. Остаток ряда координаты полюса x после восстановления по первой 21 компоненте

Спектрограмма остатка после вычитания результата восстановления ряда по 36 компонентам приведена на рис. 2.10.

Ряды dX , dY являются поправками к принятой Международным астрономическим

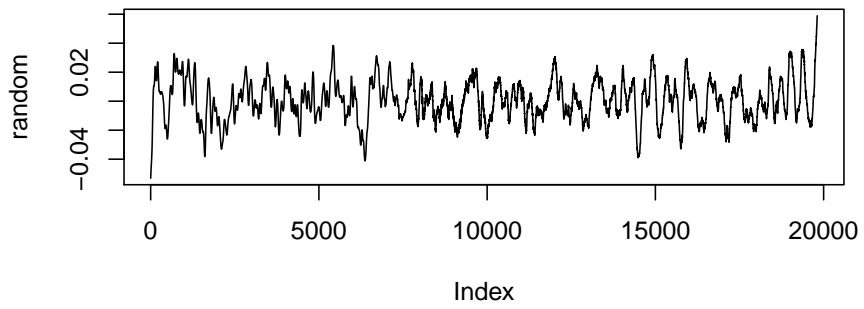


Рис. 2.6. Остаток ряда координаты полюса y после восстановления по первым 22 компонентам

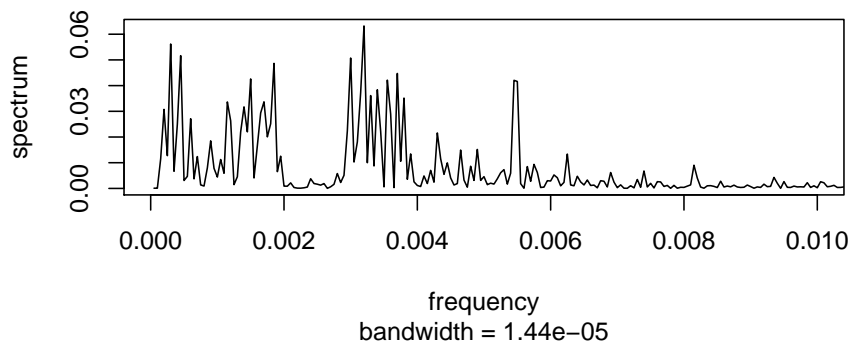


Рис. 2.7. Фрагмент спектрограммы остатка ряда координаты полюса x после выделения сигнала

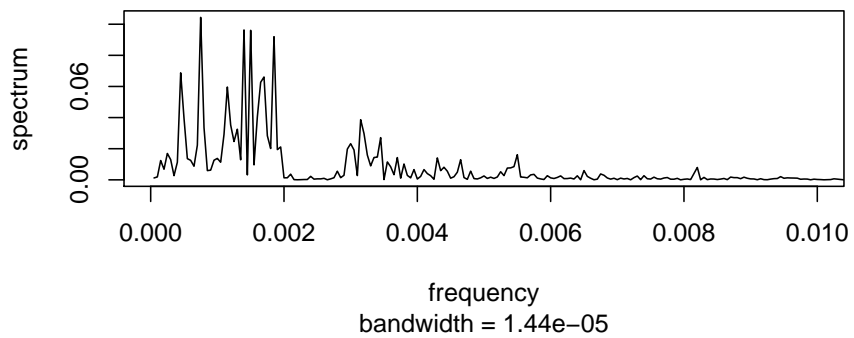


Рис. 2.8. Фрагмент спектрограммы остатка ряда координаты полюса y после выделения сигнала

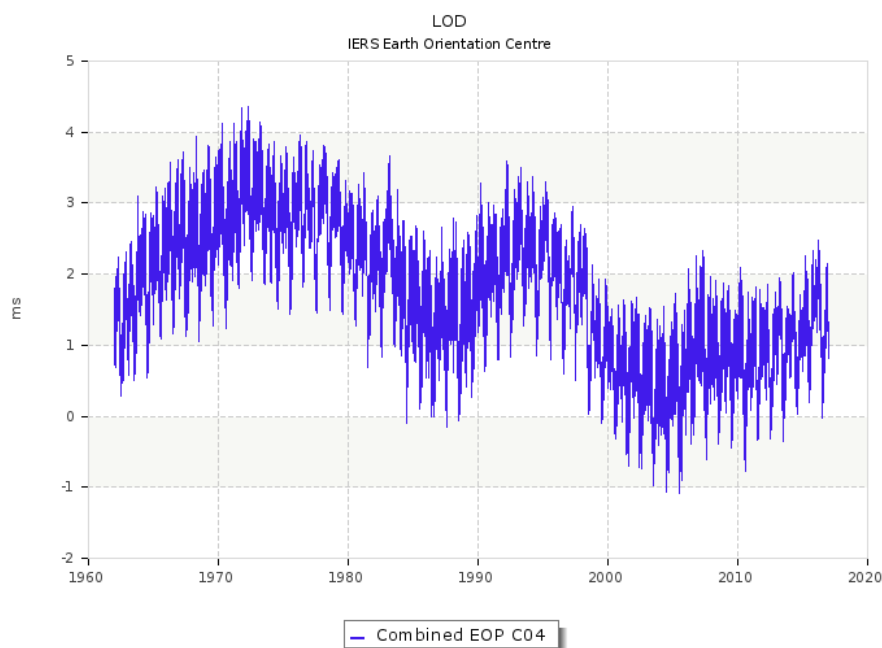


Рис. 2.9. График ряда долготы дня

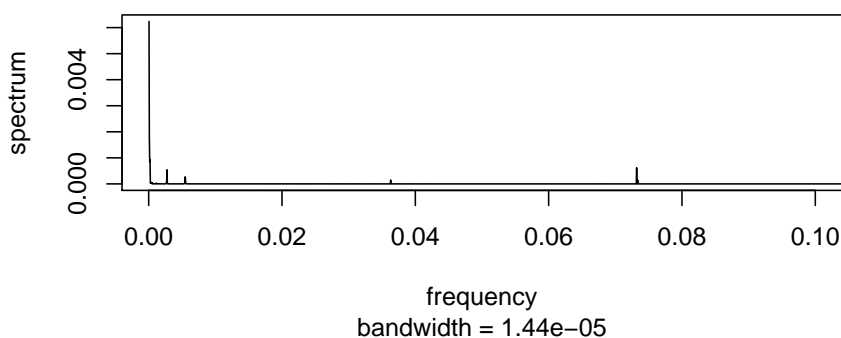


Рис. 2.10. Фрагмент спектрограммы ряда долготы дня

союзом модели прецессии-нутаии Земли, которая учитывает более тысячи колебаний. Тем не менее, в поправках по-прежнему остаются неучтённые сложно моделируемые колебания (рис. 2.13–2.14). Данные наблюдений этих рядов доступны с 1 января 1984 года. Легко заметить, что дисперсия шума, присутствующего в наблюдениях, уменьшается с течением времени. Можно предположить, что это связано с улучшением методов определения значений этих параметров в ходе технического прогресса средств измерения. Начиная с 2005 года этот шум минимален

Периоды присутствующих в рядах гармоник приблизительно равны 360, 420, 28 и 14 дням. Периодограммы этих рядов приведены на рис. 2.15–2.16. Выполним вос-

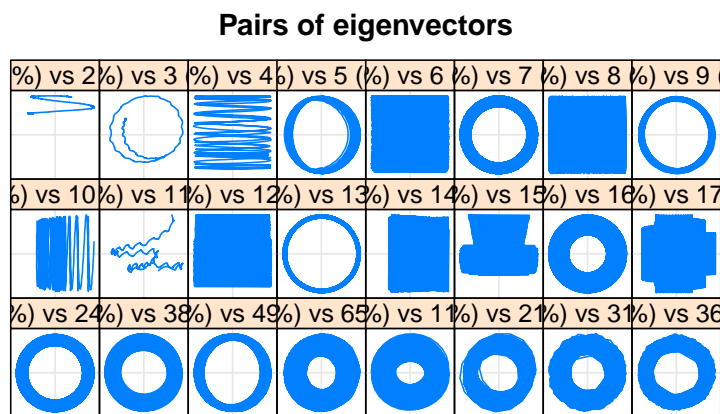


Рис. 2.11. Парные графики собственных векторов разложения ряда долготы дня

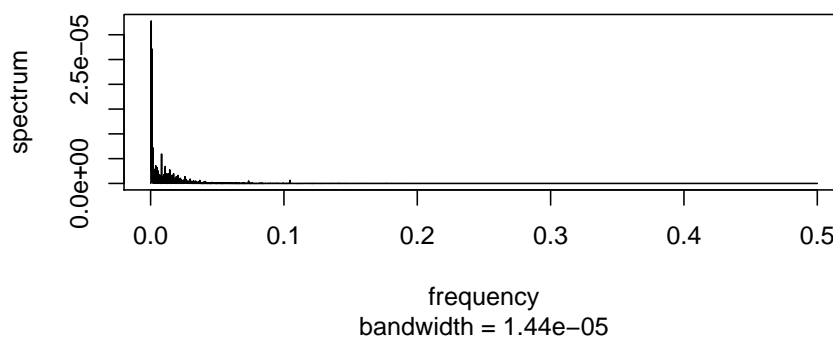


Рис. 2.12. Фрагмент спектрограммы остатков ряда долготы дня после выделения сигнала

становление этих рядов по 13 и 11 первым компонентам соответственно, используя разложения рядов с длиной окна L , равной половине их длины. Графики результатов восстановления изображены на рис. 2.17–2.18, их спектрограммы — на рис. 2.7–2.8. На графиках становится очевидной разница этих рядов, поэтому в дальнейшем также будем анализировать и прогнозировать эти ряды как самостоятельные. Спектрограммы остатков рядов dX и dY похожи на спектрограммы красного шума.

Ни в одном из анализируемых временных рядов не имеется пропусков. Дополнительные преобразований к имеющимся рядам применять не требуется.

Как правило, каждый из рассмотренных рядов может быть разложен в сумму трёх составляющих: тренд, гармонические колебания и случайный шум, причём гармоническая составляющая может быть описана моделью авторегрессии.

Метод анализа сингулярного спектра (SSA) применим для аппроксимации и про-

гнозирования широкого класса моделей рядов вида $x_n = \sum_i P_i(n)e^{\alpha_i n} \cos(2\pi\omega_i n + \varphi_i)$. В случае, когда ряды локально управляются линейно-рекуррентной формулой и в них присутствует шум, метод способен выполнять приближённое прогнозирование этих рядов. Таким образом, данный метод может быть применён для решения задачи прогнозирования рядов ПВЗ, поставленной ранее.

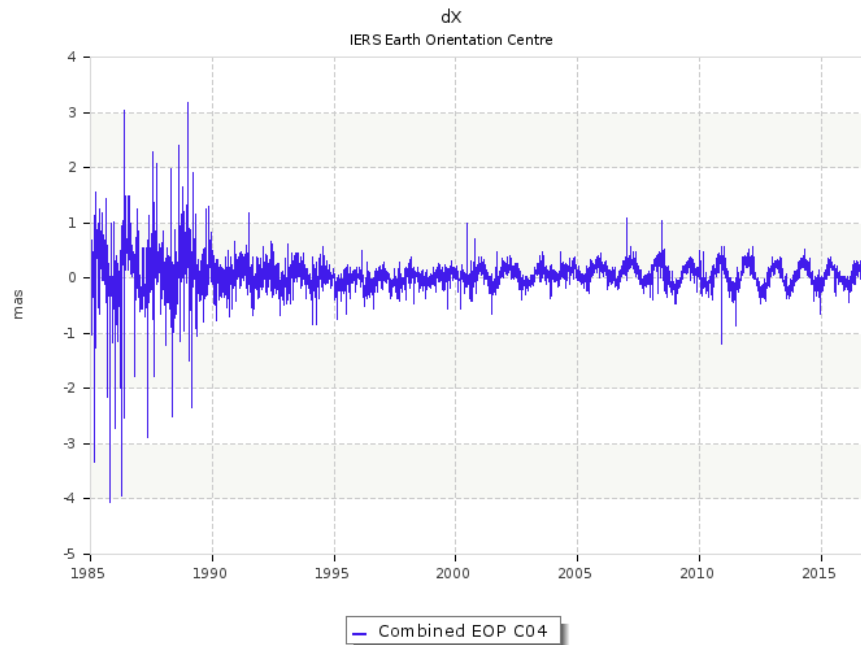


Рис. 2.13. График временного ряда поправки dX

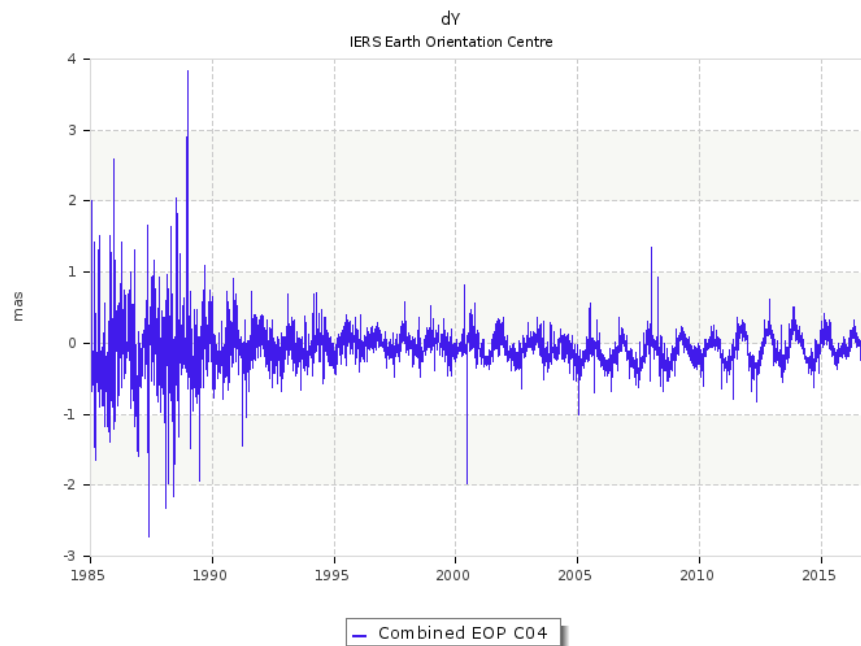


Рис. 2.14. График временного ряда поправки dY

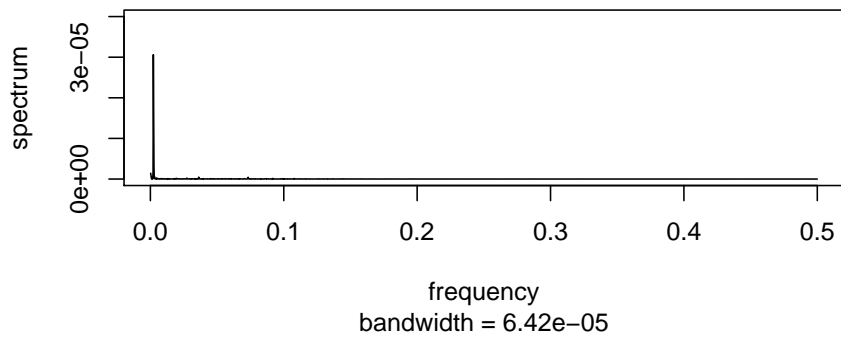


Рис. 2.15. Фрагмент спектрограммы ряда поправки dX

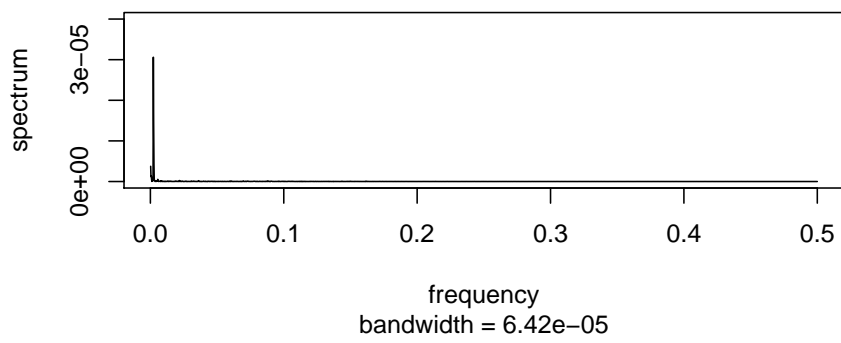


Рис. 2.16. Фрагмент спектрограммы ряда поправки dY

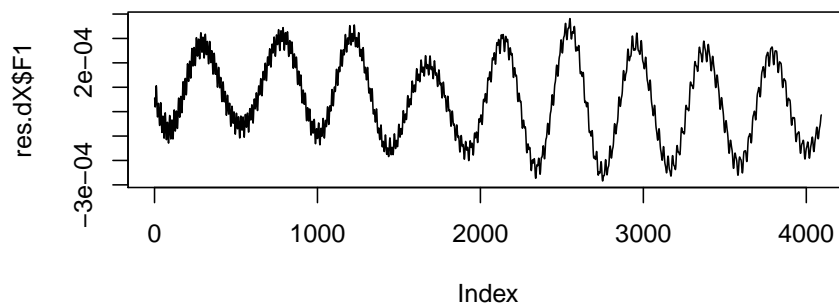


Рис. 2.17. Результат восстановления ряда поправки dX по первым 13 компонентам

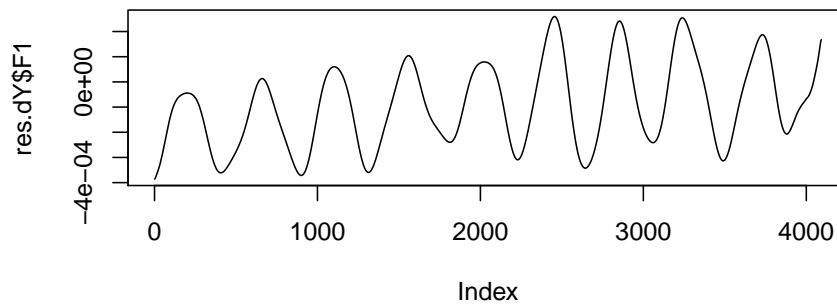


Рис. 2.18. Результат восстановления ряда поправки dY по первым 11 компонентам

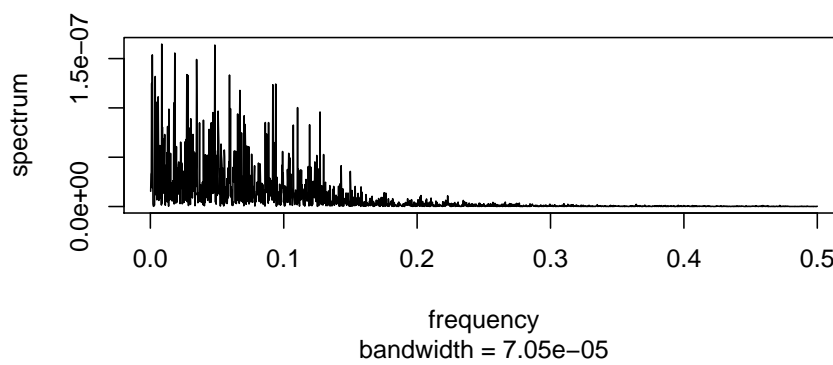


Рис. 2.19. Фрагмент спектрограммы остатков ряда поправки dX после выделения сигнала

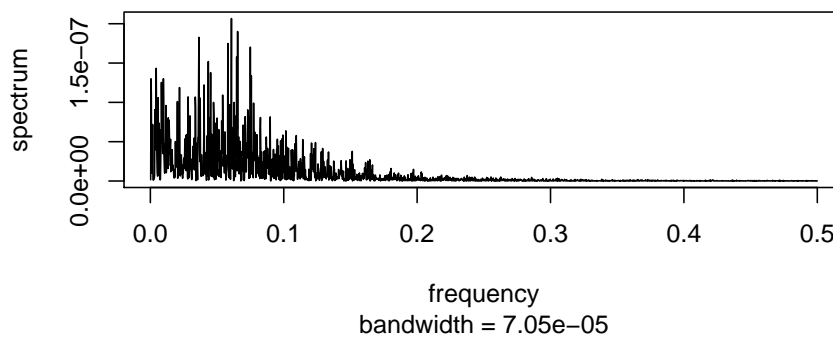


Рис. 2.20. Фрагмент спектрограммы остатков ряда поправки dY после выделения сигнала

Прогнозирование рядов ПВЗ

Пусть значения временного ряда доступны в дискретные моменты времени $t = 1, 2, \dots, N$. Обозначим временной ряд $\mathbb{X}_N = (x_1, \dots, x_N)$. В момент времени N необходимо определить значения ряда в моменты времени $N + 1, \dots, N + P$.

Обозначим $\mathbb{Y}_P = (x_{N+1}, \dots, x_{N+P})$. Величину P будем называть длиной прогноза, а \mathbb{Y}_P — прогнозом ряда \mathbb{X}_N .

Далее для удобства сравнения будем выполнять прогнозирование рядов параметров вращения Земли x, y, LOD, dX и dY на 365 дней.

3.1. Подбор параметров разложения и восстановления

Для задачи прогнозирования сигнала, содержащегося в ряде, не требуется его разделения на слагаемые, поэтому достаточно выбрать одну группировку собственных троек I_1 . Для упрощения поиска предположим, что сигналу соответствуют первые r собственных троек, упорядоченных по убыванию значения $\sqrt{\lambda_i}$, т.е. $I_1 = \{1, \dots, r\}$.

«Скользаящая» процедура, схематично изображённая на рисунке 3.1, на каждом из шагов кросс-валидации выполняет разложение ряда с фиксированной длиной окна L для каждой из n возможных группировок $I_k = \{1, \dots, k\}$, $k = 1, \dots, n$, где n — некоторое ограничение сверху на количество рассматриваемых группировок. Затем, используя разложение значений ряда на базовом отрезке, выполняет прогнозирование ряда на длину P и оценивает среднеквадратическую ошибку прогноза при сравнении с истинными значениями ряда на этом промежутке. Разложение ряда можно выполнить при помощи функции `ssa` пакета `Rssa` [7], для прогнозирования будем использовать функцию рекуррентного прогнозирования `rforecast` из того же пакета.

В качестве подходящей группировки $\{1, \dots, r^*\}$, $1 \leq r^* \leq n$, выбирается та, которая имеет наименьшую среднюю ошибку по всем шагам прогнозирования, соответствующее значение параметра L обозначим L^* . В качестве меры ошибки будем использовать среднеквадратическую ошибку MSE (mean squared error):

$$MSE_j^P(\mathbb{Y}, \hat{\mathbb{Y}}(L^*, r^*)) = \frac{1}{P} \sum_{t=1}^P (y_{j+t} - \hat{y}_{j+t}(L^*, r^*))^2,$$

где P — длина прогноза; j — момент начала прогноза; y_{j+t} — точное значение ряда \mathbb{X}_{j+t} , опубликованное в бюллетене C04; $\hat{y}_{j+t}(L^*, r^*)$ — спрогнозированное значение этого ряда, полученное с использованием разложения при выбранных значениях параметров L^* и r^* . Алгоритм вычисления MSE приведён в алгоритме 1. Псевдокод алгоритма подбора количества собственных троек для группировки с использованием скользящей процедуры приведён в алгоритме 2.

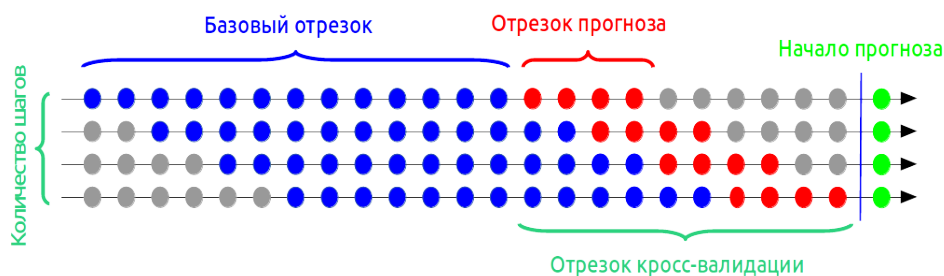


Рис. 3.1. «Скользящая» процедура подбора группировки собственных троек.

Data: \mathbb{Y}_P — точные значения ряда на сравниваемом промежутке,
 $\hat{\mathbb{Y}}_P(L^*, r^*)$ — спрогнозированные значения ряда на сравниваемом промежутке,
 P — длина прогноза,
 j — начало прогноза

Result: MSE — среднеквадратическая ошибка

```

1  $res \leftarrow 0$ ;
2 for  $t \leftarrow 1$  to  $P$  do
3    $res \leftarrow res + (y_{j+t} - \hat{y}_{j+t})^2$ ;
4 end
5 return  $\frac{res}{P}$ ;

```

Algorithm 1: Псевдокод алгоритма вычисления среднеквадратичной ошибки MSE.

Data: \mathbb{X} — временной ряд,

MJD_{start} — модифицированная юлианская дата, начиная с которой будет выполняться прогноз,

$base_len$ — длина базового отрезка,

$valid_len$ — размер отрезка кросс-валидации,

$steps$ — количество шагов кросс-валидации,

$len = P$ — длина прогноза,

$L.list$ — список значений длины окна,

n_{max} — максимальное количество компонент разложения

Result: r — выбранное число компонент

```
1  $step.len \leftarrow (valid.len - len) / (steps - 1)$ ;  
2  $L^* \leftarrow 1$ ;  $r^* \leftarrow 1$ ;  
3  $error_{min} \leftarrow \infty$ ;  
4 foreach  $L$  in  $L.list$  do  
5     for  $r \leftarrow 1$  to  $n_{max}$  do  
6          $error \leftarrow 0$ ;  
7         for  $step \leftarrow 0$  to  $steps - 1$  do  
8              $s \leftarrow ssa(\mathbb{X}[(MJD_{start} - step \times step\_len - len - base\_len) :$   
                 $(MJD_{start} - step \times step\_len - 1)], L, neig = n_{max})$ ;  
9              $f \leftarrow r forecast(s, groups = \{1, \dots, r\}, len = len, only.new = TRUE)$ ;  
10             $error \leftarrow error + MSE(\mathbb{X}[(MJD_{start} - step \times step\_len) : (MJD_{start} -$   
                 $(step - 1) \times step\_len) + 1]), f)$ ;  
11        end  
12         $mean\_error \leftarrow \frac{error}{steps}$ ;  
13        if  $mean\_error < error_{min}$  then  
14             $error_{min} \leftarrow mean\_error$ ;  
15             $L^* \leftarrow L$ ;  $r^* \leftarrow r$ ;  
16        end  
17    end  
18 end  
19 return  $L^*, r^*$ ;
```

Algorithm 2: Псевдокод алгоритма выбора параметров L^* и r^* find_L_r.

Таким образом, для получения прогноза на произвольное количество дней P следует:

1. загрузить файл `finals2000A.daily` [11] и дополнить данные из бюллетеня C04 значениями из этого файла на недостающие дни;
2. задать длину базового отрезка, на основе разложения которого будет выполняться подбор параметров и прогнозирование;
3. задать длину отрезка кросс-валидации для «скользящей» процедуры и размер шага на нём;
4. задать сетку перебираемых значений параметров L и r ;
5. на каждом шаге алгоритма для каждого заданного значения параметров L и r выполнить разложение базового отрезка и его прогнозирование на P точек по алгоритму 3;
6. оценить качество прогнозирования как MSE прогноза на тестовом отрезке;
7. выбрать значения параметров L^* и r^* , при которых ошибка прогноза минимальна.

Data: \mathbb{X} — временной ряд,

MJD_{start} — модифицированная юлианская дата, начиная с которой выполнен прогноз,

$len = P$ — длина прогноза,

$base_len$ — длина базового отрезка,

$steps$ — количество шагов скользящей процедуры;

$L.list$ — список значений длины окна;

n_{max} — максимальное количество компонент разложения

Result: \mathbb{Y} — P спрогнозированных значений ряда

- 1 $L, r \leftarrow find_L_r(\mathbb{X}, MJD_{start}, steps, len, L.list, base_len, n_{max});$ // (алгоритм 2)
- 2 $s \leftarrow ssa(\mathbb{X}[(MJD_{start} - base_len) : (MJD_{start} - 1)], L = L, neig = n_{max});$
- 3 $\mathbb{Y} \leftarrow rforecast(s, groups = I, len = len, only.new = TRUE);$
- 4 **return** $\mathbb{Y};$

Algorithm 3: Псевдокод алгоритма прогнозирования.

3.1.1. Выбор параметров L и r

В данной работе предлагается подбирать значения параметров L и r с помощью алгоритма 2 из заранее определённого множества значений. Условно разделим ряды ПВЗ на три группы и для каждой группы зададим сетки значений с фиксированным шагом. Размер шага различается между группами. Рассматриваемые значения L для различных рядов ПВЗ, разделённых на три группы, приведены в таблице 3.1.

Значения параметра r будем перебирать от 1 до 50 — для рядов x, y и LOD , от 1 до 10 — для рядов dX и dY .

Таблица 3.1. Значение параметра L для различной длины прогноза.

Параметр	L
x, y	300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900, 2100
LOD	300, 600, 900, 1200, 1500, 1800, 2100, 2400, 2700, 3000
dX, dY	250, 300, 350, 400, 450, 500

Зафиксируем длину базового отрезка (b) равной 20 годам (полагая, что каждый год длится 365 дней, т.е. длина базового отрезка составляет 7300 дней), длину отрезка валидации (v) — 5 лет, размер шага по отрезку валидации — 365 дней. Выполним прогнозирование каждого из рядов ПВЗ на 365 дней для первых тридцати последовательных дней и для 50 дней, выбранных с шагом в 7 дней начиная с 1 января, в каждом из перечисленных годов: 2011, 2012, 2013, 2014 и 2015. Рассмотрим значения параметров L^* и r^* , выбранные алгоритмом 2, при которых достигается минимальная среднеквадратичная ошибка каждого из прогнозов $MSE_{min}^P(j) = \min_{L,r} MSE^P(j, L, r)$, где j — день начала прогноза. По всем наборам параметров вычислим длину максимальной непрерывной последовательности выбора одного и того же значения, среднее, моду и медиану значений параметров (приложение А). Средние значения параметров L и r по двум упомянутым последовательностям дней приведены в таблицах 3.2–3.6.

Из приведённых таблиц можно сделать вывод, что при прогнозировании на последовательном участке используемый алгоритм выбирает одинаковые значения параметров L и r (как правило, не менее 15 дней подряд), что и следовало ожидать в силу схожего поведения каждого из рядов на некотором отрезке времени (здесь — 30 дней). При этом значения параметра L для рядов dX, dY иногда достигают крайнего

Таблица 3.2. Средние значения параметров для ряда x .

	2006	2007	2008	2009	2010
Среднее L по 30 последовательным дням	1100	1153	1060	900	900
Среднее L по 50 дням с шагом в 7 дней	1060	1128	956	1024	1100
Среднее r по 30 последовательным дням	26	23	27	20	15
Среднее r по 50 дням с шагом в 7 дней	23	24	22	19	20

Таблица 3.3. Средние значения параметров для ряда y .

	2006	2007	2008	2009	2010
Среднее L по 30 последовательным дням	580	953	900	1300	966
Среднее L по 50 дням с шагом в 7 дней	944	1200	1084	1124	928
Среднее r по 30 последовательным дням	17	23	18	23	25
Среднее r по 50 дням с шагом в 7 дней	19	23	22	21	19

Таблица 3.4. Средние значения параметров для ряда LOD .

	2006	2007	2008	2009	2010
Среднее L по 30 последовательным дням	600	600	760	790	600
Среднее L по 50 дням с шагом в 7 дней	636	612	714	810	816
Среднее r по 30 последовательным дням	12	10	15	22	25
Среднее r по 50 дням с шагом в 7 дней	15	15	19	22	23

Таблица 3.5. Средние значения параметров для ряда dX .

	2006	2007	2008	2009	2010
Среднее L по 30 последовательным дням	475	450	465	500	450
Среднее L по 50 дням с шагом в 7 дней	481	466	488	462	439
Среднее r по 30 последовательным дням	3	3	3	3	3
Среднее r по 50 дням с шагом в 7 дней	3	3	3	3	3

из рассмотренных значений в 500 точек. Для остальных рядов значение длины окна L оказывается одним из промежуточных.

При этом значения параметра r не превышают 30 для рядов x, y, LOD и 5 для рядов dX, dY , следовательно, перебор значений параметра r может быть сокращён в целях ускорения работы алгоритма прогнозирования.

Таблица 3.6. Средние значения параметров для ряда dY .

	2006	2007	2008	2009	2010
Среднее L по 30 последовательным дням	466	500	450	495	495
Среднее L по 50 дням с шагом в 7 дней	482	468	454	470	409
Среднее r по 30 последовательным дням	4	4	3	4	5
Среднее r по 50 дням с шагом в 7 дней	4	4	3	4	4

3.1.2. Выбор длины базового отрезка

В различных работах используются различные длины базового отрезка, на основе которого выполняется прогнозирование: 5, 10 и 15 лет [14, 15], причём ограничения сверху на длину используемого отрезка накладываемся в силу трудоёмкости используемых методов.

Сравним результаты прогнозирования по базовому отрезку длиной 5, 10, 15 и 20 лет, зафиксировав размер отрезка валидации $v_0 = 5$ (лет), а количество шагов кросс-валидации — 10.

Случаи, при которых в процессе перебора длина окна L оказывается больше длины базового отрезка, не рассматриваются.

Для сравнения качества прогнозирования при различных значениях параметров предлагается следующая процедура. В каждом году из пяти предложенных (2006, 2007, 2008, 2009, 2010) выбирается по 50 дней, начиная с 1 января каждого года с шагом в 7 дней (всего 250 дней). Для каждого из выбранных дней выполняется прогноз пяти рядов ПВЗ на 365 дней. Для каждого из полученных прогнозов будет вычислена среднеквадратическая ошибка прогноза, полученного при фиксированном значении v_0 для каждого из рассматриваемых значений b_i , при этом параметры L и r подбираются так, как было описано выше в алгоритме 2: $MSE_j^P(b_i, v_0) = \min_{L,r} MSE(j, P, L, r, b_i, v_0)$.

В пределах каждого года будет вычислено среднее значение MSE по этому году и интервал, в пределах которого оказываются 80% полученных значений MSE. Наилучшим будет считаться тот набор параметров, для которого среднее значение MSE_{mean}^P по 250 дням окажется наименьшей: $b^* = \underset{b_i}{\operatorname{argmin}} MSE_{mean}^P(b_i, v_0)$.

Так как значения рядов dX и dY в бюллетене C04 приведены начиная с 1 января 1984 года, то суммарная длина базового отрезка и отрезка кросс-валидации для этих рядов не должна превышать 22 лет, поэтому значение базового отрезка в 20 лет для

этих рядов рассматриваться не будут.

Далее в таблицах 3.7–3.10 приводятся результаты сравнения прогнозов, выполненных при различных значениях длины базового участка. Ячейки, содержащие минимальные значения среднеквадратической ошибки среди рассматриваемых прогнозов, выделены цветом.

Из анализа таблиц следует, что наименьшая в среднем среднеквадратическая ошибка прогноза для всех рядов x, y, LOD, dX, dY получается при длине базового отрезка в 15 лет.

3.1.3. Выбор длины отрезка кросс-валидации

Зафиксируем выбранное на предыдущем шаге значения длины базового отрезка $b^* = 15$ (лет) и количество шагов кросс-валидации (10).

Рассмотрим следующие возможные значения длины отрезка кросс-валидации: 3, 5, 7, 10 лет для рядов x, y, LOD и 3, 5, 7 лет для рядов dX и dY . При выборе наборов значений учитывалось, что суммарная длина базового отрезка и отрезка кросс-валидации для рядов dX, dY не превышала 22 лет.

Аналогично предыдущему шагу, будем выбирать значение параметра v , минимизирующее ошибку прогноза $MSE_j^P(b^*, v_i) = \min_{L,r} MSE(j, P, L, r, b^*, v_i)$ при фиксированном b^* и для всех рассматриваемых значений v_i : $v^* = \operatorname{argmin}_{v_i} MSE_{mean}^P(b^*, v_i)$. Значения параметров L и r всё так же будут подбираться по алгоритму 2.

Результаты сравнения прогнозов, полученных при изменении параметра длины отрезка кросс-валидации, приведены в таблицах 3.12–3.16.

Минимальные средние значения ошибки для рядов x и y достигаются при длине отрезка кросс-валидации в 7 лет. Для ряда LOD предлагается использовать длину отрезка валидации в 10 лет. Для рядов dX, dY предлагается использовать значение в 7 лет.

Таким образом, для выполнения прогнозирования рядов ПВЗ на 365 дней предлагается использовать значения параметров, приведённые в таблице 3.22. Параметры L и r будут выбираться по алгоритму 2: параметр L — среди значений, приведённых в таблице 3.1, а r — с учётом ограничений, принятых в разделе 3.1.1.

Графики прогнозов, полученных на 365 дней, начиная с 1 января 2015 (MJD = 57023), изображены на рисунках 3.2–3.6. Для прогнозирования были использованы

значения параметров, приведённые в таблице 3.22, оптимальные значения L^* и r^* подобраны автоматически по алгоритму 2 и приведены в таблице 3.23. Графики прогнозов для некоторых других начальных дат приведены в приложении Б.

Выполним разложение и восстановление базовых участков рядов ПВЗ, соответствующих 1 января 2015 (MJD = 57023), с использованием значений параметров, приведённых в таблице 3.23 (рис. 3.7–3.11). Из графиков видно, что восстановления рядов x, y выполняется довольно близко к исходным значениям. Ряд LOD восстанавливается не так точно, но это объясняется известным фактом о сложной структуре и непостоянстве ряда. Восстановление рядов dX, dY по подобранному числу компонент позволяет избавиться от шума, присутствующего в исходных рядах.

3.2. Сравнение результатов прогнозирования

Для прогнозирования рядов ПВЗ с использованием рекуррентного алгоритма прогнозирования метода анализа сингулярного спектра будем использовать наилучшие для каждого из рядов значения параметров, упомянутые выше. Выполним сравнение полученных результатов прогнозирования по аналогичной процедуре, получив прогнозы для 50 дней в каждом году из пяти перечисленных.

Будем сравнивать результаты работы предлагаемого алгоритма с прогнозами МСВЗ (публикуемыми в бюллетене А) и Пулковской обсерватории [16]. Значения в файлах прогнозов Пулковской обсерватории необходимо предварительно домножить на соответствующие множители, чтобы данные измерялись в шкале, используемой МСВЗ. Упомянутые прогнозы составляются на 365 дней: бюллетень А — еженедельно начиная с 6 января 2005 года, прогнозы Пулковской обсерватории — ежедневно с 27 августа 2010 года.

Прогнозы ПВЗ на следующие 365 дней публикуются МСВЗ в бюллетене А еженедельно, по четвергам. По адресу <https://datacenter.iers.org/web/guest/bulletins/-/somos/5Rgv/version/6> находится архив всех опубликованных бюллетеней в различных форматах. Данные, на основе которых был выполнен последний опубликованный прогноз, находятся по адресу <http://maia.usno.navy.mil/bulla-data.html>.

Бюллетень А содержит оценки значений ПВЗ за последние 7 дней, прогнозы рядов на следующие 365 дней и коэффициенты для аналитической формулы, позволяющей пользователям самостоятельно получать оценки ПВЗ на произвольное количество дней

вперёд. Бюллетень А не содержит прогноза для рядов dX , dY и LOD .

Сравнение прогнозов из различных источников приведено в таблицах 3.17–3.21. Из приведённых таблиц следует, что в среднем прогнозы рядов x , y , LOD , dY , полученные с использованием метода анализа сингулярного спектра на 365 дней, оказываются лучше, чем аналогичные прогнозы, публикуемые Пулковской обсерваторией и МСВЗ, а качество прогноз ряда dX незначительно уступает прогнозу Пулковской обсерватории.

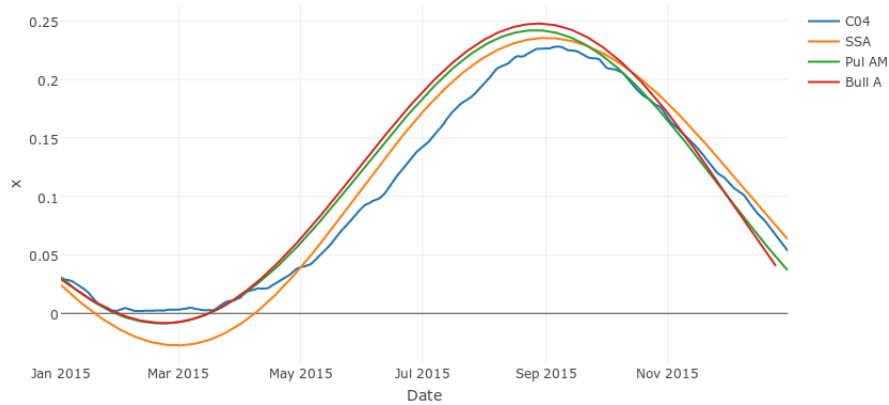


Рис. 3.2. График прогнозов ряда x на 365 дней для $MJD = 57023$.

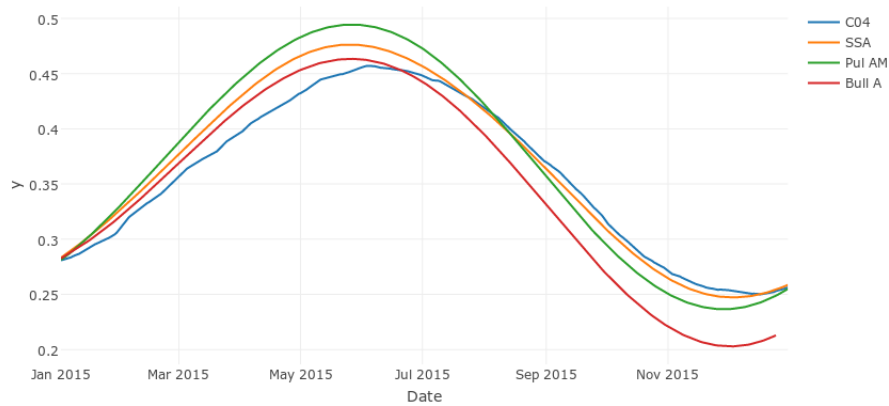


Рис. 3.3. График прогнозов ряда y на 365 дней для $MJD = 57023$.

3.3. Описание интерфейса веб-приложения

Описанный выше алгоритм прогнозирования был реализован на языке **R** с использованием библиотеки `Rssa` [7]. Созданный скрипт выполняет загрузку последнего выпуска бюллетеня `C04` [9] и дополняет его значения рядов x , y , LOD , dX и dY `rapid`-данными из файла `finals2000A.daily` [11] до текущего дня. Затем для подбора оптимальных зна-

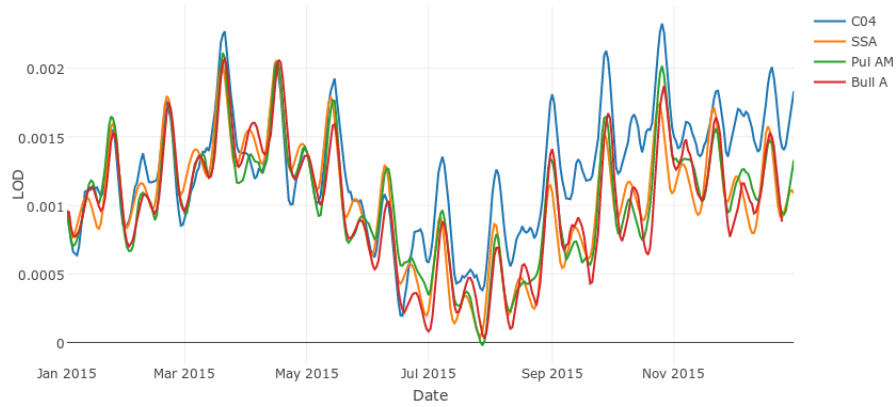


Рис. 3.4. График прогнозов ряда LOD на 365 дней для $MJD = 57023$.

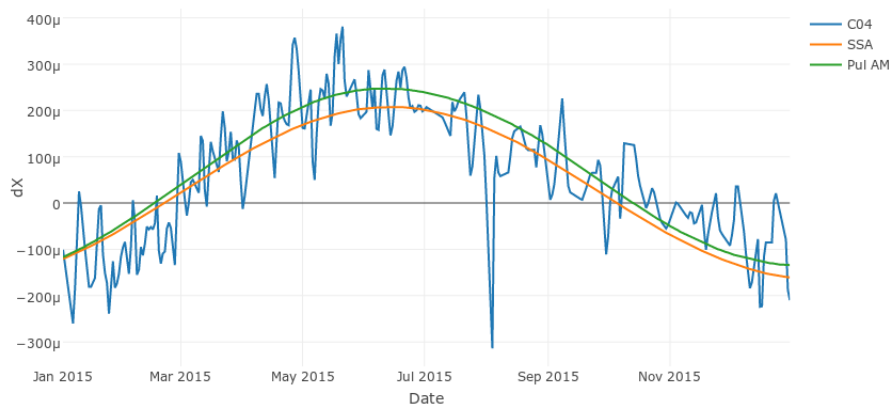


Рис. 3.5. График прогнозов ряда dX на 365 дней для $MJD = 57023$.

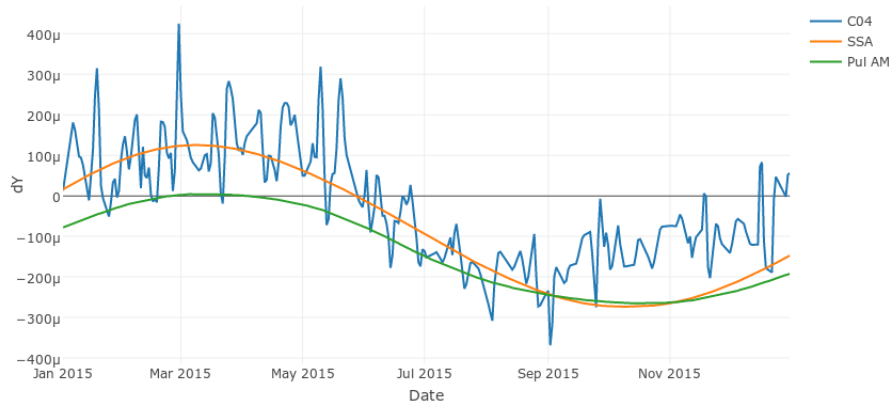


Рис. 3.6. График прогнозов ряда dY на 365 дней для $MJD = 57023$.

чений параметров L^* и r^* — длины окна разложения и количества первых собственных троек в группировке соответственно — каждого из пяти временных рядов ПВЗ запускается алгоритм 2 с использованием параметров кросс-валидации, приведённых в таблице 3.22. После того, как значения L^* и r^* будут получены, выполняется прогнозирование

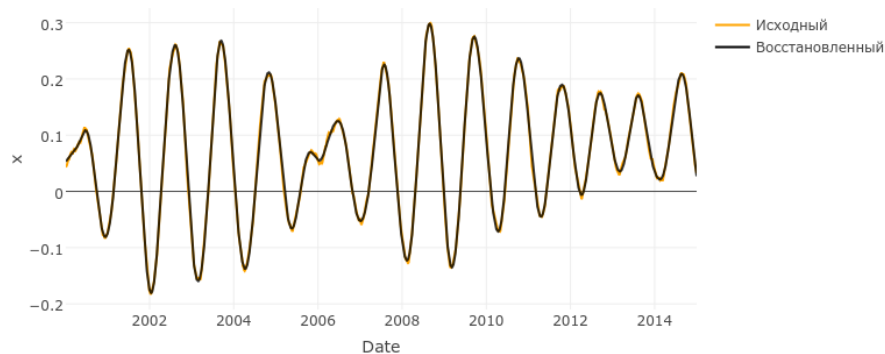


Рис. 3.7. График восстановления базового участка ряда x для $MJD = 57023$.

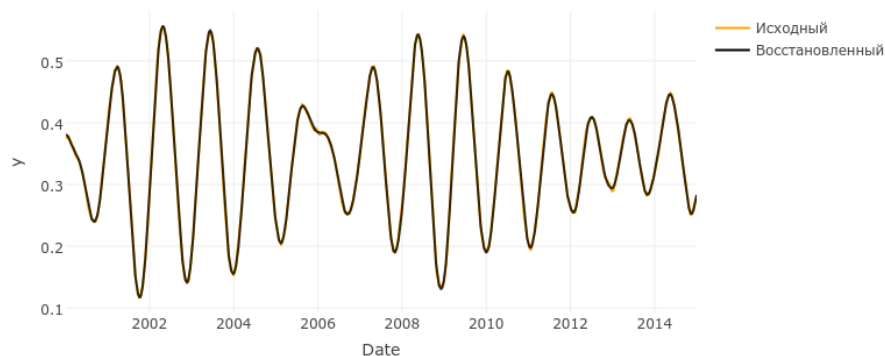


Рис. 3.8. График восстановления базового участка ряда y для $MJD = 57023$.

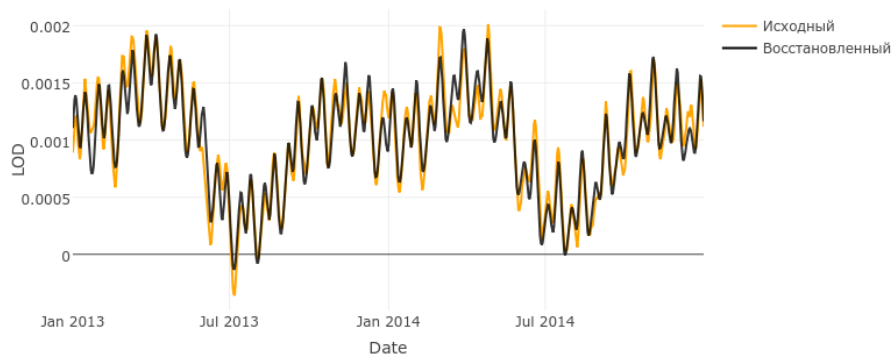


Рис. 3.9. График фрагмента восстановления базового участка ряда LOD для $MJD = 57023$.

рядов на указанное количество дней по алгоритму 3.

Для публикации результатов работы алгоритма при помощи пакета RShiny было создано веб-приложение «SSA EOP Forecast» (доступно по адресу <http://eoppredict.ru/>), которое выполняет прогнозирование рядов ПВЗ на 365 дней и отображает графики полученных прогнозов и параметры, при которых они были получены.

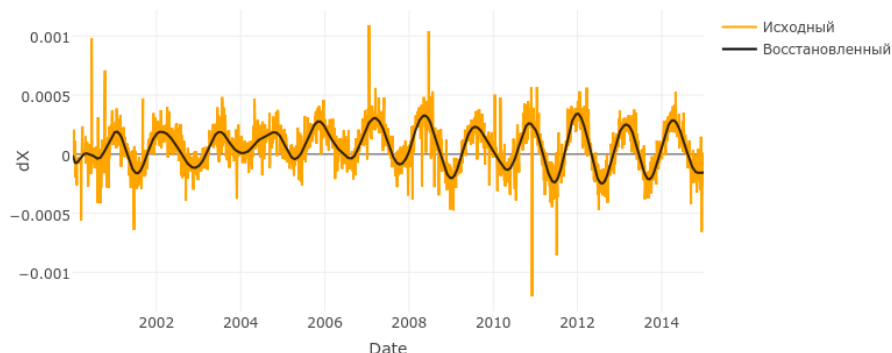


Рис. 3.10. График восстановления базового участка ряда dX для $MJD = 57023$.

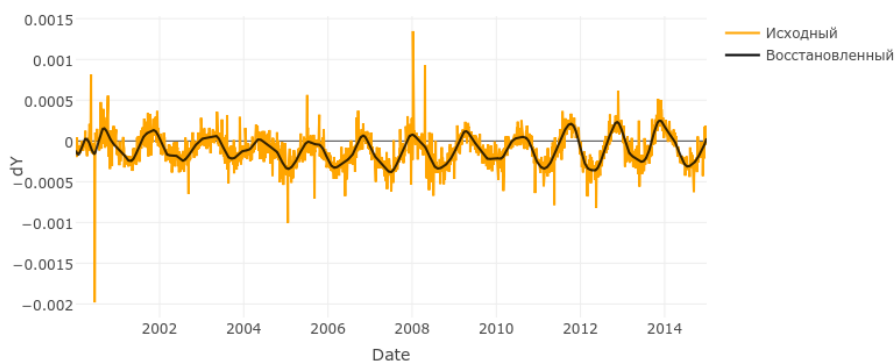


Рис. 3.11. График восстановления базового участка ряда dY для $MJD = 57023$.

3.3.1. Страница «Forecast for Today»

Прогнозы рядов x, y, LOD, dX, dY , полученные с помощью предложенного в данной работе алгоритма, отображаются в виде графиков, начиная с текущего дня. На графиках рядов x, y, LOD дополнительно приводится прогноз МСВЗ из бюллетеня А. Модифицированная юлианская дата (MJD) текущего дня отображается в дополнительной панели слева. Пользователю предоставляется возможность переключать подписи временной оси графиков между датами и MJD. График каждого ряда ПВЗ сопровождается значениями параметров L^* и r^* , при которых был получен этот прогноз, а также графики значений MSE для каждой рассмотренной пары значений параметров L и r .

Полученные при помощи метода «Гусеница»-SSA прогнозы доступны для скачивания в виде файлов при нажатии на кнопку «Download».

3.3.2. Страница «Generate Forecast»

Пользователю предоставляется возможность выполнить прогноз самостоятельно, используя информацию о среднеквадратической ошибке, получаемой при различных

значениях параметров L и r , на произвольное количество дней, а также загрузить полученный им прогноз в виде файла. Для этого необходимо выбрать ряд, для которого выполняется прогноз, задать параметр длины окна L и количество первых собственных троек разложения r , по которым будет выполняться аппроксимация сигнала. Дополнительно требуется указать длину прогноза в днях и длину базового отрезка, по которому будет строиться прогноз.

После нажатия на кнопку «Generate» результаты прогнозирования ряда будут отображены в виде графика, а в боковой панели слева появится кнопка «Download» для загрузки файла с прогнозом.

3.3.3. Страница «Compare Forecasts»

На отдельной вкладке предоставляется интерфейс для изображения графиков прогнозов ПВЗ из различных источников на указанный день в прошлом и сравнения среднеквадратической ошибки каждого из прогнозов.

Каждый из прогнозов, изображённых на графике, может быть скрыт или снова показан нажатием на соответствующий пункт в легенде. Кроме того, графики могут быть приближены для более подробного рассмотрения. В начале страницы находится сводная таблица среднеквадратических ошибок каждого из прогнозов с возможностью сортировки по каждому из рядов ПВЗ.

Бюллетень А не содержит прогнозов рядов dX, dY , поэтому эти ячейки таблицы остаются пустыми, а графики этих рядов не содержат соответствующих бюллетеню А линий.

Кнопка «Download» позволяет скачать прогноз на выбранную дату, полученный с помощью разработанного в данной работе алгоритма.

3.3.4. Страница «About»

Последняя вкладка содержит краткую информацию о методе и пакетах, которые были использованы для реализации данного приложения. Внизу страницы приводятся ссылки на релевантные источники информации и на репозиторий на сайте <http://github.com>, содержащий исходный код приложения.

Таблица 3.7. MSE прогнозов при различной длине базового отрезка для ряда x .

	5 лет	80% интервал	10 лет	80% интервал	15 лет	80% интервал	20 лет	80% интервал
2006	8.2×10^{-3}	$[9.8 \times 10^{-4}; 1.5 \times 10^{-2}]$	1.7×10^{-3}	$[5.9 \times 10^{-4}; 2.7 \times 10^{-3}]$	1.2×10^{-3}	$[4.2 \times 10^{-4}; 2.1 \times 10^{-3}]$	2.0×10^{-3}	$[4.7 \times 10^{-4}; 4.0 \times 10^{-3}]$
2007	2.0×10^{-3}	$[3.8 \times 10^{-4}; 4.8 \times 10^{-3}]$	7.8×10^{-4}	$[2.5 \times 10^{-4}; 1.7 \times 10^{-3}]$	4.3×10^{-4}	$[2.0 \times 10^{-4}; 6.5 \times 10^{-4}]$	8.7×10^{-4}	$[1.5 \times 10^{-4}; 2.5 \times 10^{-3}]$
2008	1.3×10^{-3}	$[8.2 \times 10^{-5}; 3.0 \times 10^{-3}]$	8.4×10^{-4}	$[4.1 \times 10^{-4}; 1.4 \times 10^{-3}]$	9.6×10^{-4}	$[4.9 \times 10^{-4}; 1.3 \times 10^{-3}]$	9.7×10^{-4}	$[6.1 \times 10^{-4}; 1.3 \times 10^{-3}]$
2009	5.0×10^{-4}	$[1.3 \times 10^{-4}; 9.4 \times 10^{-4}]$	3.9×10^{-4}	$[1.2 \times 10^{-4}; 9.7 \times 10^{-4}]$	6.6×10^{-4}	$[2.9 \times 10^{-4}; 1.4 \times 10^{-3}]$	6.1×10^{-4}	$[2.2 \times 10^{-4}; 1.4 \times 10^{-3}]$
2010	1.2×10^{-3}	$[3.3 \times 10^{-4}; 2.8 \times 10^{-3}]$	1.0×10^{-3}	$[4.3 \times 10^{-4}; 2.6 \times 10^{-3}]$	9.6×10^{-4}	$[4.4 \times 10^{-4}; 1.8 \times 10^{-3}]$	1.1×10^{-3}	$[5.8 \times 10^{-4}; 2.1 \times 10^{-3}]$
Среднее	2.6×10^{-3}		9.5×10^{-4}		8.5×10^{-4}		1.1×10^{-3}	
Медиана	9.9×10^{-4}		7.0×10^{-4}		6.1×10^{-4}		8.2×10^{-4}	

Таблица 3.8. MSE прогнозов при различной длине базового отрезка для ряда y .

	5 лет	80% интервал	10 лет	80% интервал	15 лет	80% интервал	20 лет	80% интервал
2006	4.0×10^{-3}	$[3.3 \times 10^{-4}; 7.4 \times 10^{-3}]$	2.4×10^{-3}	$[5.2 \times 10^{-4}; 4.8 \times 10^{-3}]$	1.4×10^{-3}	$[5.8 \times 10^{-4}; 2.4 \times 10^{-3}]$	2.0×10^{-3}	$[7.9 \times 10^{-4}; 3.8 \times 10^{-3}]$
2007	1.7×10^{-3}	$[4.7 \times 10^{-4}; 3.9 \times 10^{-3}]$	8.0×10^{-4}	$[2.4 \times 10^{-4}; 1.4 \times 10^{-3}]$	4.4×10^{-4}	$[1.7 \times 10^{-4}; 9.7 \times 10^{-4}]$	3.4×10^{-4}	$[1.3 \times 10^{-4}; 4.8 \times 10^{-4}]$
2008	2.2×10^{-3}	$[3.4 \times 10^{-4}; 5.3 \times 10^{-3}]$	1.2×10^{-3}	$[5.1 \times 10^{-4}; 3.0 \times 10^{-3}]$	8.5×10^{-4}	$[2.1 \times 10^{-4}; 1.5 \times 10^{-3}]$	9.3×10^{-4}	$[3.7 \times 10^{-4}; 1.6 \times 10^{-3}]$
2009	8.6×10^{-4}	$[1.9 \times 10^{-4}; 1.8 \times 10^{-3}]$	7.1×10^{-4}	$[4.8 \times 10^{-5}; 1.5 \times 10^{-3}]$	7.9×10^{-4}	$[5.0 \times 10^{-5}; 2.6 \times 10^{-3}]$	9.9×10^{-4}	$[4.7 \times 10^{-5}; 2.3 \times 10^{-3}]$
2010	8.4×10^{-4}	$[1.5 \times 10^{-4}; 1.7 \times 10^{-3}]$	1.1×10^{-3}	$[4.1 \times 10^{-4}; 1.5 \times 10^{-3}]$	1.0×10^{-3}	$[4.4 \times 10^{-4}; 1.6 \times 10^{-3}]$	1.4×10^{-3}	$[6.0 \times 10^{-4}; 2.0 \times 10^{-3}]$
Среднее	1.9×10^{-3}		1.2×10^{-3}		9.0×10^{-4}		1.1×10^{-3}	
Медиана	1.1×10^{-3}		8.9×10^{-4}		6.8×10^{-4}		9.0×10^{-4}	

Таблица 3.9. MSE прогнозов при различной длине базового отрезка для ряда LOD .

	5 лет	80% интервал	10 лет	80% интервал	15 лет	80% интервал	20 лет	80% интервал
2006	3.4×10^{-7}	$[1.5 \times 10^{-7}; 6.0 \times 10^{-7}]$	3.5×10^{-7}	$[1.4 \times 10^{-7}; 4.7 \times 10^{-7}]$	1.8×10^{-7}	$[7.8 \times 10^{-8}; 3.1 \times 10^{-7}]$	1.6×10^{-7}	$[7.4 \times 10^{-8}; 2.9 \times 10^{-7}]$
2007	5.6×10^{-7}	$[1.1 \times 10^{-7}; 9.3 \times 10^{-7}]$	3.7×10^{-7}	$[6.8 \times 10^{-8}; 8.4 \times 10^{-7}]$	1.0×10^{-7}	$[5.1 \times 10^{-8}; 2.1 \times 10^{-7}]$	1.1×10^{-7}	$[8.1 \times 10^{-8}; 1.7 \times 10^{-7}]$
2008	6.2×10^{-8}	$[3.9 \times 10^{-8}; 8.4 \times 10^{-8}]$	7.0×10^{-8}	$[4.4 \times 10^{-8}; 8.9 \times 10^{-8}]$	6.0×10^{-8}	$[4.2 \times 10^{-8}; 8.2 \times 10^{-8}]$	7.1×10^{-8}	$[4.7 \times 10^{-8}; 1.1 \times 10^{-7}]$
2009	1.1×10^{-7}	$[7.4 \times 10^{-8}; 1.5 \times 10^{-7}]$	7.9×10^{-8}	$[4.5 \times 10^{-8}; 1.2 \times 10^{-7}]$	7.8×10^{-8}	$[5.2 \times 10^{-8}; 9.6 \times 10^{-8}]$	7.2×10^{-8}	$[4.3 \times 10^{-8}; 1.1 \times 10^{-7}]$
2010	1.1×10^{-7}	$[5.2 \times 10^{-8}; 2.0 \times 10^{-7}]$	8.5×10^{-8}	$[4.6 \times 10^{-8}; 1.2 \times 10^{-7}]$	9.0×10^{-8}	$[6.9 \times 10^{-8}; 1.3 \times 10^{-7}]$	8.6×10^{-8}	$[5.3 \times 10^{-8}; 1.6 \times 10^{-7}]$
Среднее	2.4×10^{-7}		1.9×10^{-7}		1.0×10^{-7}		1.0×10^{-7}	
Медиана	1.4×10^{-7}		9.8×10^{-8}		8.1×10^{-8}		8.5×10^{-8}	

Таблица 3.10. MSE прогнозов при различной длине базового отрезка для ряда dX .

	5 лет	80% интервал	10 лет	80% интервал	15 лет	80% интервал
2006	1.34×10^{-8}	$[1.1 \times 10^{-8}; 1.6 \times 10^{-8}]$	1.14×10^{-8}	$[1.1 \times 10^{-8}; 1.2 \times 10^{-8}]$	1.18×10^{-8}	$[1.1 \times 10^{-8}; 1.3 \times 10^{-8}]$
2007	1.35×10^{-8}	$[9.2 \times 10^{-9}; 1.9 \times 10^{-8}]$	1.37×10^{-8}	$[9.1 \times 10^{-9}; 2.0 \times 10^{-8}]$	1.40×10^{-8}	$[1.0 \times 10^{-8}; 1.8 \times 10^{-8}]$
2008	1.65×10^{-8}	$[1.4 \times 10^{-8}; 2.0 \times 10^{-8}]$	1.91×10^{-8}	$[1.4 \times 10^{-8}; 2.4 \times 10^{-8}]$	1.84×10^{-8}	$[1.4 \times 10^{-8}; 2.3 \times 10^{-8}]$
2009	2.25×10^{-8}	$[1.9 \times 10^{-8}; 2.8 \times 10^{-8}]$	1.86×10^{-8}	$[1.4 \times 10^{-8}; 2.1 \times 10^{-8}]$	1.78×10^{-8}	$[1.4 \times 10^{-8}; 2.0 \times 10^{-8}]$
2010	3.45×10^{-8}	$[3.1 \times 10^{-8}; 4.0 \times 10^{-8}]$	3.38×10^{-8}	$[3.0 \times 10^{-8}; 3.6 \times 10^{-8}]$	3.19×10^{-8}	$[3.0 \times 10^{-8}; 3.4 \times 10^{-8}]$
Среднее	2.01×10^{-8}		1.93×10^{-8}		1.88×10^{-8}	
Медиана	1.83×10^{-8}		1.77×10^{-8}		1.75×10^{-8}	

Таблица 3.11. MSE прогнозов при различной длине базового отрезка для ряда dY .

	5 лет	80% интервал	10 лет	80% интервал	15 лет	80% интервал
2006	1.35×10^{-8}	$[9.0 \times 10^{-9}; 2.0 \times 10^{-8}]$	1.17×10^{-8}	$[8.9 \times 10^{-9}; 1.5 \times 10^{-8}]$	9.51×10^{-9}	$[8.1 \times 10^{-9}; 1.1 \times 10^{-8}]$
2007	2.65×10^{-8}	$[2.0 \times 10^{-8}; 3.1 \times 10^{-8}]$	2.60×10^{-8}	$[2.2 \times 10^{-8}; 3.1 \times 10^{-8}]$	2.45×10^{-8}	$[2.0 \times 10^{-8}; 2.9 \times 10^{-8}]$
2008	1.31×10^{-8}	$[7.8 \times 10^{-9}; 2.0 \times 10^{-8}]$	1.16×10^{-8}	$[7.3 \times 10^{-9}; 1.7 \times 10^{-8}]$	1.12×10^{-8}	$[7.2 \times 10^{-9}; 1.6 \times 10^{-8}]$
2009	1.16×10^{-8}	$[1.0 \times 10^{-8}; 1.3 \times 10^{-8}]$	1.17×10^{-8}	$[1.0 \times 10^{-8}; 1.3 \times 10^{-8}]$	1.07×10^{-8}	$[9.5 \times 10^{-9}; 1.2 \times 10^{-8}]$
2010	1.68×10^{-8}	$[1.2 \times 10^{-8}; 2.3 \times 10^{-8}]$	1.56×10^{-8}	$[1.1 \times 10^{-8}; 2.3 \times 10^{-8}]$	1.49×10^{-8}	$[1.1 \times 10^{-8}; 2.2 \times 10^{-8}]$
Среднее	1.63×10^{-8}		1.53×10^{-8}		1.42×10^{-8}	
Медиана	1.37×10^{-8}		1.26×10^{-8}		1.13×10^{-8}	

Таблица 3.12. MSE прогнозов при различной длине отрезка кросс-валидации для ряда x .

	3 года	80% интервал	5 лет	80% интервал	7 лет	80% интервал	10 лет	80% интервал
2006	1.2×10^{-3}	$[3.9 \times 10^{-4}; 2.1 \times 10^{-3}]$	1.2×10^{-3}	$[4.2 \times 10^{-4}; 2.1 \times 10^{-3}]$	1.2×10^{-3}	$[5.4 \times 10^{-4}; 1.9 \times 10^{-3}]$	1.3×10^{-3}	$[4.5 \times 10^{-4}; 1.9 \times 10^{-3}]$
2007	1.1×10^{-3}	$[4.3 \times 10^{-4}; 1.7 \times 10^{-3}]$	4.3×10^{-4}	$[2.0 \times 10^{-4}; 6.5 \times 10^{-4}]$	4.1×10^{-4}	$[2.4 \times 10^{-4}; 6.3 \times 10^{-4}]$	5.9×10^{-4}	$[2.7 \times 10^{-4}; 1.0 \times 10^{-3}]$
2008	1.2×10^{-3}	$[6.9 \times 10^{-4}; 1.6 \times 10^{-3}]$	9.6×10^{-4}	$[4.9 \times 10^{-4}; 1.3 \times 10^{-3}]$	8.9×10^{-4}	$[5.1 \times 10^{-4}; 1.4 \times 10^{-3}]$	9.5×10^{-4}	$[4.1 \times 10^{-4}; 1.6 \times 10^{-3}]$
2009	6.3×10^{-4}	$[3.9 \times 10^{-4}; 1.4 \times 10^{-3}]$	6.6×10^{-4}	$[2.9 \times 10^{-4}; 1.4 \times 10^{-3}]$	5.4×10^{-4}	$[2.4 \times 10^{-4}; 9.8 \times 10^{-4}]$	6.1×10^{-4}	$[3.5 \times 10^{-4}; 1.1 \times 10^{-3}]$
2010	1.1×10^{-3}	$[5.2 \times 10^{-4}; 1.6 \times 10^{-3}]$	9.6×10^{-4}	$[4.4 \times 10^{-4}; 1.8 \times 10^{-3}]$	1.1×10^{-3}	$[5.5 \times 10^{-4}; 2.2 \times 10^{-3}]$	1.1×10^{-3}	$[5.0 \times 10^{-4}; 2.0 \times 10^{-3}]$
Среднее	1.0×10^{-3}		8.5×10^{-4}		8.3×10^{-4}		9.1×10^{-4}	
Медиана	9.6×10^{-4}		6.1×10^{-4}		6.3×10^{-4}		7.1×10^{-4}	

Таблица 3.13. MSE прогнозов при различной длине отрезка кросс-валидации для ряда y .

	3 года	80% интервал	5 лет	80% интервал	7 лет	80% интервал	10 лет	80% интервал
2006	1.1×10^{-3}	$[5.5 \times 10^{-4}; 1.7 \times 10^{-3}]$	1.4×10^{-3}	$[5.8 \times 10^{-4}; 2.4 \times 10^{-3}]$	1.3×10^{-3}	$[6.5 \times 10^{-4}; 1.9 \times 10^{-3}]$	1.3×10^{-3}	$[6.3 \times 10^{-4}; 2.5 \times 10^{-3}]$
2007	5.6×10^{-4}	$[2.1 \times 10^{-4}; 1.0 \times 10^{-3}]$	4.4×10^{-4}	$[1.7 \times 10^{-4}; 9.7 \times 10^{-4}]$	4.4×10^{-4}	$[1.7 \times 10^{-4}; 8.9 \times 10^{-4}]$	4.4×10^{-4}	$[1.4 \times 10^{-4}; 8.1 \times 10^{-4}]$
2008	8.3×10^{-4}	$[3.6 \times 10^{-4}; 1.8 \times 10^{-3}]$	8.5×10^{-4}	$[2.1 \times 10^{-4}; 1.5 \times 10^{-3}]$	7.5×10^{-4}	$[2.7 \times 10^{-4}; 1.4 \times 10^{-3}]$	6.8×10^{-4}	$[3.6 \times 10^{-4}; 1.4 \times 10^{-3}]$
2009	8.2×10^{-4}	$[6.4 \times 10^{-5}; 1.9 \times 10^{-3}]$	7.9×10^{-4}	$[5.0 \times 10^{-5}; 2.6 \times 10^{-3}]$	9.0×10^{-4}	$[4.3 \times 10^{-5}; 2.9 \times 10^{-3}]$	7.9×10^{-4}	$[5.6 \times 10^{-5}; 2.6 \times 10^{-3}]$
2010	1.2×10^{-3}	$[7.2 \times 10^{-4}; 1.9 \times 10^{-3}]$	1.0×10^{-3}	$[4.4 \times 10^{-4}; 1.6 \times 10^{-3}]$	1.0×10^{-3}	$[6.3 \times 10^{-4}; 1.3 \times 10^{-3}]$	1.2×10^{-3}	$[3.0 \times 10^{-4}; 2.0 \times 10^{-3}]$
Среднее	9.2×10^{-4}		9.0×10^{-4}		8.8×10^{-4}		8.8×10^{-4}	
Медиана	7.8×10^{-4}		6.8×10^{-4}		7.6×10^{-4}		7.1×10^{-4}	

Таблица 3.14. MSE прогнозов при различной длине отрезка кросс-валидации для ряда LOD .

	3 года	80% интервал	5 лет	80% интервал	7 лет	80% интервал	10 лет	80% интервал
2006	1.9×10^{-7}	$[8.0 \times 10^{-8}; 3.3 \times 10^{-7}]$	1.8×10^{-7}	$[7.8 \times 10^{-8}; 3.1 \times 10^{-7}]$	1.8×10^{-7}	$[7.9 \times 10^{-8}; 3.1 \times 10^{-7}]$	1.8×10^{-7}	$[6.7 \times 10^{-8}; 3.5 \times 10^{-7}]$
2007	1.2×10^{-7}	$[5.4 \times 10^{-8}; 2.1 \times 10^{-7}]$	1.0×10^{-7}	$[5.1 \times 10^{-8}; 2.1 \times 10^{-7}]$	1.0×10^{-7}	$[5.0 \times 10^{-8}; 2.1 \times 10^{-7}]$	9.8×10^{-8}	$[4.2 \times 10^{-8}; 2.8 \times 10^{-7}]$
2008	5.7×10^{-8}	$[4.5 \times 10^{-8}; 7.7 \times 10^{-8}]$	6.0×10^{-8}	$[4.2 \times 10^{-8}; 8.2 \times 10^{-8}]$	6.1×10^{-8}	$[4.4 \times 10^{-8}; 8.0 \times 10^{-8}]$	5.5×10^{-8}	$[3.9 \times 10^{-8}; 8.7 \times 10^{-8}]$
2009	8.2×10^{-8}	$[5.4 \times 10^{-8}; 1.1 \times 10^{-7}]$	7.8×10^{-8}	$[5.2 \times 10^{-8}; 9.6 \times 10^{-8}]$	7.4×10^{-8}	$[5.1 \times 10^{-8}; 8.9 \times 10^{-8}]$	7.2×10^{-8}	$[5.0 \times 10^{-8}; 8.7 \times 10^{-8}]$
2010	8.2×10^{-8}	$[4.7 \times 10^{-8}; 1.1 \times 10^{-7}]$	9.0×10^{-8}	$[6.9 \times 10^{-8}; 1.3 \times 10^{-7}]$	9.8×10^{-8}	$[6.0 \times 10^{-8}; 1.3 \times 10^{-7}]$	8.5×10^{-8}	$[6.1 \times 10^{-8}; 1.0 \times 10^{-7}]$
Среднее	1.1×10^{-7}		1.0×10^{-7}		1.0×10^{-7}		9.8×10^{-8}	
Медиана	8.1×10^{-8}		8.1×10^{-8}		8.1×10^{-8}		7.4×10^{-8}	

Таблица 3.15. MSE прогнозов при различной длине отрезка кросс-валидации для ряда dX .

	3 года	80% интервал	5 лет	80% интервал	7 лет	80% интервал
2006	1.17×10^{-8}	$[1.1 \times 10^{-8}; 1.3 \times 10^{-8}]$	1.18×10^{-8}	$[1.1 \times 10^{-8}; 1.3 \times 10^{-8}]$	1.26×10^{-8}	$[1.1 \times 10^{-8}; 1.6 \times 10^{-8}]$
2007	1.42×10^{-8}	$[1.0 \times 10^{-8}; 1.8 \times 10^{-8}]$	1.40×10^{-8}	$[1.0 \times 10^{-8}; 1.8 \times 10^{-8}]$	1.38×10^{-8}	$[9.8 \times 10^{-9}; 1.8 \times 10^{-8}]$
2008	1.81×10^{-8}	$[1.3 \times 10^{-8}; 2.3 \times 10^{-8}]$	1.84×10^{-8}	$[1.4 \times 10^{-8}; 2.3 \times 10^{-8}]$	1.83×10^{-8}	$[1.4 \times 10^{-8}; 2.3 \times 10^{-8}]$
2009	1.80×10^{-8}	$[1.3 \times 10^{-8}; 2.1 \times 10^{-8}]$	1.78×10^{-8}	$[1.4 \times 10^{-8}; 2.0 \times 10^{-8}]$	1.78×10^{-8}	$[1.3 \times 10^{-8}; 2.0 \times 10^{-8}]$
2010	3.38×10^{-8}	$[3.2 \times 10^{-8}; 3.6 \times 10^{-8}]$	3.19×10^{-8}	$[3.0 \times 10^{-8}; 3.4 \times 10^{-8}]$	3.14×10^{-8}	$[3.0 \times 10^{-8}; 3.4 \times 10^{-8}]$
Среднее	1.91×10^{-8}		1.88×10^{-8}		1.88×10^{-8}	
Медиана	1.72×10^{-8}		1.75×10^{-8}		1.75×10^{-8}	

Таблица 3.16. MSE прогнозов при различной длине отрезка кросс-валидации для ряда dY .

	3 года	80% интервал	5 лет	80% интервал	7 лет	80% интервал
2006	9.79×10^{-9}	$[8.0 \times 10^{-9}; 1.1 \times 10^{-8}]$	9.51×10^{-9}	$[8.1 \times 10^{-9}; 1.1 \times 10^{-8}]$	9.34×10^{-9}	$[8.3 \times 10^{-9}; 1.0 \times 10^{-8}]$
2007	2.43×10^{-8}	$[1.9 \times 10^{-8}; 2.9 \times 10^{-8}]$	2.45×10^{-8}	$[2.0 \times 10^{-8}; 2.9 \times 10^{-8}]$	2.42×10^{-8}	$[1.9 \times 10^{-8}; 2.8 \times 10^{-8}]$
2008	1.11×10^{-8}	$[7.1 \times 10^{-9}; 1.6 \times 10^{-8}]$	1.12×10^{-8}	$[7.2 \times 10^{-9}; 1.6 \times 10^{-8}]$	1.11×10^{-8}	$[7.1 \times 10^{-9}; 1.6 \times 10^{-8}]$
2009	1.02×10^{-8}	$[8.9 \times 10^{-9}; 1.1 \times 10^{-8}]$	1.07×10^{-8}	$[9.5 \times 10^{-9}; 1.2 \times 10^{-8}]$	1.09×10^{-8}	$[9.7 \times 10^{-9}; 1.3 \times 10^{-8}]$
2010	1.56×10^{-8}	$[1.2 \times 10^{-8}; 2.2 \times 10^{-8}]$	1.49×10^{-8}	$[1.1 \times 10^{-8}; 2.2 \times 10^{-8}]$	1.41×10^{-8}	$[1.0 \times 10^{-8}; 2.0 \times 10^{-8}]$
Среднее	1.42×10^{-8}		1.42×10^{-8}		1.39×10^{-8}	
Медиана	1.12×10^{-8}		1.13×10^{-8}		1.10×10^{-8}	

Таблица 3.17. MSE прогнозов ряда x из различных источников

	SSA	80% интервал	Pulkovo AM	80% интервал	Bulletin A	80% интервал
2011	6.7×10^{-4}	$[2.9 \times 10^{-4}; 1.1 \times 10^{-3}]$	8.4×10^{-4}	$[3.1 \times 10^{-4}; 1.3 \times 10^{-3}]$	1.2×10^{-3}	$[4.0 \times 10^{-4}; 1.7 \times 10^{-3}]$
2012	8.7×10^{-4}	$[1.6 \times 10^{-4}; 1.8 \times 10^{-3}]$	7.4×10^{-4}	$[3.6 \times 10^{-4}; 1.4 \times 10^{-3}]$	4.9×10^{-4}	$[2.6 \times 10^{-4}; 7.6 \times 10^{-4}]$
2013	9.4×10^{-4}	$[3.2 \times 10^{-4}; 1.6 \times 10^{-3}]$	1.4×10^{-3}	$[6.3 \times 10^{-4}; 2.2 \times 10^{-3}]$	7.9×10^{-4}	$[5.1 \times 10^{-4}; 1.1 \times 10^{-3}]$
2014	4.4×10^{-4}	$[9.0 \times 10^{-5}; 8.2 \times 10^{-4}]$	8.8×10^{-4}	$[4.1 \times 10^{-4}; 1.3 \times 10^{-3}]$	7.2×10^{-4}	$[2.7 \times 10^{-4}; 1.0 \times 10^{-3}]$
2015	6.7×10^{-4}	$[2.9 \times 10^{-4}; 1.3 \times 10^{-3}]$	4.5×10^{-4}	$[1.8 \times 10^{-4}; 7.3 \times 10^{-4}]$	4.8×10^{-4}	$[1.3 \times 10^{-4}; 8.3 \times 10^{-4}]$
Среднее	7.2×10^{-4}		8.6×10^{-4}		7.5×10^{-4}	
Медиана	5.9×10^{-4}		7.2×10^{-4}		7.0×10^{-4}	

Таблица 3.18. MSE прогнозов ряда y из различных источников

	SSA	80% интервал	Pulkovo AM	80% интервал	Bulletin A	80% интервал
2011	3.7×10^{-4}	$[1.0 \times 10^{-4}; 6.5 \times 10^{-4}]$	1.1×10^{-3}	$[5.0 \times 10^{-4}; 2.1 \times 10^{-3}]$	7.7×10^{-4}	$[3.6 \times 10^{-4}; 1.1 \times 10^{-3}]$
2012	7.3×10^{-4}	$[2.7 \times 10^{-4}; 1.3 \times 10^{-3}]$	2.7×10^{-4}	$[1.2 \times 10^{-4}; 4.0 \times 10^{-4}]$	4.9×10^{-4}	$[2.5 \times 10^{-4}; 7.0 \times 10^{-4}]$
2013	7.7×10^{-4}	$[2.4 \times 10^{-4}; 1.3 \times 10^{-3}]$	7.7×10^{-4}	$[3.5 \times 10^{-4}; 9.7 \times 10^{-4}]$	1.1×10^{-3}	$[4.1 \times 10^{-4}; 1.7 \times 10^{-3}]$
2014	3.3×10^{-4}	$[6.0 \times 10^{-5}; 5.8 \times 10^{-4}]$	8.4×10^{-4}	$[3.4 \times 10^{-4}; 1.3 \times 10^{-3}]$	8.3×10^{-4}	$[2.6 \times 10^{-4}; 1.4 \times 10^{-3}]$
2015	8.6×10^{-4}	$[1.8 \times 10^{-4}; 1.6 \times 10^{-3}]$	8.0×10^{-4}	$[3.7 \times 10^{-4}; 1.5 \times 10^{-3}]$	1.1×10^{-3}	$[3.4 \times 10^{-4}; 2.0 \times 10^{-3}]$
Среднее	6.1×10^{-4}		7.6×10^{-4}		8.5×10^{-4}	
Медиана	4.9×10^{-4}		6.3×10^{-4}		7.7×10^{-4}	

Таблица 3.19. MSE прогнозов ряда LOD из различных источников

	SSA	80% интервал	Пулково АМ	80% интервал
2011	8.7×10^{-8}	$[6.1 \times 10^{-8}; 1.3 \times 10^{-7}]$	1.3×10^{-7}	$[9.4 \times 10^{-8}; 1.7 \times 10^{-7}]$
2012	9.4×10^{-8}	$[6.1 \times 10^{-8}; 1.3 \times 10^{-7}]$	8.2×10^{-8}	$[5.8 \times 10^{-8}; 1.0 \times 10^{-7}]$
2013	4.9×10^{-8}	$[3.2 \times 10^{-8}; 7.6 \times 10^{-8}]$	5.4×10^{-8}	$[4.3 \times 10^{-8}; 6.5 \times 10^{-8}]$
2014	5.2×10^{-8}	$[2.3 \times 10^{-8}; 1.1 \times 10^{-7}]$	5.5×10^{-8}	$[3.5 \times 10^{-8}; 9.0 \times 10^{-8}]$
2015	1.7×10^{-7}	$[1.0 \times 10^{-7}; 2.4 \times 10^{-7}]$	1.8×10^{-7}	$[9.0 \times 10^{-8}; 2.5 \times 10^{-7}]$
Среднее	9.1×10^{-8}		1.0×10^{-7}	
Медиана	7.8×10^{-8}		8.2×10^{-8}	

Таблица 3.20. MSE прогнозов ряда dX из различных источников

	SSA	80% интервал	Пулково АМ	80% интервал
2011	2.1×10^{-8}	$[1.6 \times 10^{-8}; 2.5 \times 10^{-8}]$	1.7×10^{-8}	$[9.2 \times 10^{-9}; 2.8 \times 10^{-8}]$
2012	1.0×10^{-8}	$[7.3 \times 10^{-9}; 1.3 \times 10^{-8}]$	1.1×10^{-8}	$[8.2 \times 10^{-9}; 1.3 \times 10^{-8}]$
2013	1.4×10^{-8}	$[1.0 \times 10^{-8}; 1.6 \times 10^{-8}]$	8.1×10^{-9}	$[6.1 \times 10^{-9}; 1.3 \times 10^{-8}]$
2014	8.8×10^{-9}	$[8.3 \times 10^{-9}; 9.2 \times 10^{-9}]$	1.3×10^{-8}	$[8.9 \times 10^{-9}; 1.7 \times 10^{-8}]$
2015	9.6×10^{-9}	$[6.3 \times 10^{-9}; 1.2 \times 10^{-8}]$	8.2×10^{-9}	$[6.7 \times 10^{-9}; 1.0 \times 10^{-8}]$
Среднее	1.3×10^{-8}		1.1×10^{-8}	
Медиана	1.1×10^{-8}		1.0×10^{-8}	

Таблица 3.21. MSE прогнозов ряда dY из различных источников

	SSA	80% интервал	Pulkovo AM	80% интервал
2011	1.7×10^{-8}	$[1.5 \times 10^{-8}; 2.0 \times 10^{-8}]$	3.5×10^{-8}	$[1.2 \times 10^{-8}; 5.0 \times 10^{-8}]$
2012	2.0×10^{-8}	$[1.4 \times 10^{-8}; 3.0 \times 10^{-8}]$	2.1×10^{-8}	$[1.2 \times 10^{-8}; 3.0 \times 10^{-8}]$
2013	1.9×10^{-8}	$[9.7 \times 10^{-9}; 3.4 \times 10^{-8}]$	2.8×10^{-8}	$[1.9 \times 10^{-8}; 3.5 \times 10^{-8}]$
2014	1.4×10^{-8}	$[1.1 \times 10^{-8}; 1.8 \times 10^{-8}]$	1.2×10^{-8}	$[7.5 \times 10^{-9}; 1.7 \times 10^{-8}]$
2015	1.1×10^{-8}	$[8.8 \times 10^{-9}; 1.3 \times 10^{-8}]$	1.5×10^{-8}	$[1.0 \times 10^{-8}; 1.9 \times 10^{-8}]$
Среднее	1.6×10^{-8}		2.2×10^{-8}	
Медиана	1.5×10^{-8}		1.9×10^{-8}	

Таблица 3.22. Выбранные значения параметров для прогнозирования рядов ПВЗ на 365 дней.

Параметр	x	y	LOD	dX	dY
b^* (лет)	15	15	15	15	15
v^* (лет)	7	7	10	7	7

Таблица 3.23. Выбранные значения параметров для прогнозирования рядов ПВЗ на 365 дней с 1 января 2015 года (MJD = 57023).

Параметр	x	y	LOD	dX	dY
L^*	700	500	900	350	300
r^*	10	10	19	4	5

Заключение

В ходе работы алгоритм рекуррентного прогнозирования метода «Гусеница»-SSA был применён для прогнозирования временных рядов пяти параметров вращения Земли: x , y , LOD , dX и dY . Прогнозы рядов довольно хорошо согласуются с точными значениями на рассмотренном промежутке времени.

В настоящей работе были получены следующие результаты:

- изучены источники и организован автоматический сбор данных, необходимых для анализа и прогнозирования временных рядов ПВЗ;
- разработана методика, с помощью которой был выполнен подбор оптимальных параметров алгоритма рекуррентного прогнозирования метода «Гусеница»-SSA для прогнозирования рядов ПВЗ;
- реализован на языке R алгоритм автоматического подбора параметров для ежедневного прогнозирования пяти рядов ПВЗ при помощи метода «Гусеница»-SSA;
- выполнено сравнение прогнозов, полученных с использованием автоматически подобранных параметров, с прогнозами, предоставляемыми еженедельно МСВЗ и ежедневно Пулковской обсерваторией;
- создано веб-приложение, с помощью которого результаты работы программы прогнозирования публикуются ежедневно, а также выполняется сравнение прогнозов из различных источников с точными значениями.

Сравнение результатов прогнозирования рядов ПВЗ, полученных с помощью алгоритма рекуррентного прогнозирования метода «Гусеница»-SSA, с их точными значениями из бюллетеня С04 позволяет сделать вывод о том, что данный метод действительно применим к решению задачи прогнозирования этих рядов.

При сравнении прогнозов, полученных с использованием разработанной методики, оказалось, что в среднем полученный результат оказывается лучше прогнозов для рядов x , y , LOD и dY , с которыми выполнялось сравнение, а ошибка прогноза ряда dX оказывается незначительно больше, чем ошибка прогнозов Пулковской обсерватории.

Список литературы

1. Earth Orientation Parameters Prediction Comparison Campaign. — URL: http://users.cbk.waw.pl/~kalma/EOP_PCC/.
2. Earth Orientation Parameters Combination of Prediction Pilot Project. — URL: <http://maia.usno.navy.mil/eopcppp/contributions.html>.
3. Миллер Н. О. Прогнозирование движения полюса с помощью ССА // Известия ГАО РАН. — 2016. — Т. 223. — С. 119–123.
4. Исследование геодинамических рядов методом главных компонент / В. Л. Горшков, Н. О. Миллер, Н. Р. Персиянинова, Е. Я. Прудникова // Известия ГАО РАН. — 2000. — Т. 214. — С. 173–180.
5. Витязев В. В., Миллер Н. О., Прудникова Е. Я. Использование сингулярного спектрального анализа при исследовании движения полюса // Вестник Санкт-Петербургского университета. Серия 1. Математика. Механика. Астрономия. — 2012. — № 2.
6. Multivariate and 2D extensions of singular spectrum analysis with the Rssa package / N. Golyandina, A. Korobeynikov, A. Shlemov, K. Usevich // Journal of Statistical Software. — 2015. — Vol. 67. — P. 1–78.
7. Korobeynikov Anton, Shlemov Alex, Usevich Konstantin, Golyandina Nina. — Rssa: A Collection of Methods for Singular Spectrum Analysis, 2015. — R package version 0.13-1. URL: <https://cran.r-project.org/package=Rssa>.
8. Golyandina N., Korobeynikov A. Basic singular spectrum analysis and forecasting with R // Computational Statistics & Data Analysis. — 2014. — Vol. 71. — P. 934–954.
9. International Earth rotation and reference systems service Earth orientation parameters EOP (IERS) 14 C04. — URL: https://hpiers.obspm.fr/iers/eop/eopc04/eopc04_IAU2000.62-now.
10. The combined Earth Orientation solution C04. — URL: https://hpiers.obspm.fr/eoppp/eop/eopc04_05/C04_guide.pdf.
11. Daily Rapid EOP Data for the last 90 days (IAU2000). — URL: <https://datacenter.iers.org/eop/-/somos/5Rgv/latestXL/13/finals2000A.daily>.
12. Андерсон Т. Статистический анализ временных рядов. — Москва : Мир, 1976.
13. Кузнецов В. В. Физика Земли. — Новосибирск, 2011.

14. Malkin Z. Employing combination procedures to short-time EOP prediction // Artificial Satellites. — 2010. — Vol. 45, no. 2. — P. 87–93.
15. Kosek W. Future improvements in EOP prediction // Geodesy for Planet Earth. — Springer, 2012. — P. 513–520.
16. Pulkovo EOP and Reference Systems Analysis Center (PERSAC). — URL: <http://www.gao.spb.ru/english/as/persac/eopcphp/>.

Приложение А

Таблицы значений параметров L^* и r^*

Таблица А.1. Значение параметра L^* для ряда x на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	30: $L^* = 1100$	26: $L^* = 1100$	22: $L^* = 1500$	30: $L^* = 900$	30: $L^* = 900$
Минимальное	1100	1100	900	900	900
Максимальное	1100	1500	1500	900	900
Среднее	1100	1153	1060	900	900
Мода	1100	1100	900	900	900
Медиана	1100	1100	900	900	900

Таблица А.2. Значение параметра L^* для ряда y на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	18: $L^* = 500$	26: $L^* = 900$	30: $L^* = 900$	30: $L^* = 1300$	20: $L^* = 1100$
Минимальное	500	900	900	1300	900
Максимальное	700	1300	900	1300	1100
Среднее	580	953	900	1300	966
Мода	500	900	900	1300	900
Медиана	500	900	900	1300	900

Таблица А.3. Значение параметра L^* для ряда LOD на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	30: $L^* = 600$	30: $L^* = 600$	16: $L^* = 900$	19: $L^* = 900$	30: $L^* = 600$
Минимальное	600	600	600	600	600
Максимальное	600	600	900	900	600
Среднее	600	600	760	790	600
Мода	600	600	900	900	600
Медиана	600	600	900	900	600

Таблица А.4. Значение параметра L^* для ряда dX на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	15: $L^* = 500$	30: $L^* = 450$	9: $L^* = 450$	30: $L^* = 500$	30: $L^* = 450$
Минимальное	450	450	400	500	450
Максимальное	500	450	500	500	450
Среднее	475	450	465	500	450
Мода	500	450	450	500	450
Медиана	475	450	450	500	450

Таблица А.5. Значение параметра L^* для ряда dY на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	13: $L^* = 500$	30: $L^* = 500$	30: $L^* = 450$	27: $L^* = 450$	29: $L^* = 500$
Минимальное	400	500	450	450	350
Максимальное	500	500	450	500	500
Среднее	466	500	450	495	495
Мода	500	500	450	500	500
Медиана	500	500	450	500	500

Таблица А.6. Значение параметра L^* для ряда x на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	17: $L^* = 1100$	15: $L^* = 1100$	27: $L^* = 900$	22: $L^* = 900$	11: $L^* = 900$
Минимальное	500	900	900	900	700
Максимальное	2100	2100	1500	2100	2100
Среднее	1060	1128	956	1024	1100
Мода	1100	1100	900	900	900
Медиана	1100	1100	900	900	900

Таблица А.7. Значение параметра L^* для ряда y на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	7: $L^* = 900$	13: $L^* = 1300$	12: $L^* = 900$	10: $L^* = 1300$	28: $L^* = 900$
Минимальное	500	500	900	900	900
Максимальное	1500	1500	1300	1500	1300
Среднее	944	1200	1084	1124	928
Мода	500	1300	900	900	900
Медиана	900	1300	900	1100	900

Таблица А.8. Значение параметра L^* для ряда LOD на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	33: $L^* = 600$	28: $L^* = 600$	18: $L^* = 900$	16: $L^* = 900$	19: $L^* = 600$
Минимальное	600	600	600	600	600
Максимальное	900	900	900	900	900
Среднее	636	612	714	810	816
Мода	600	600	600	900	900
Медиана	600	600	600	900	900

Таблица А.9. Значение параметра L^* для ряда dX на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	16: $L^* = 450$	8: $L^* = 450$	32: $L^* = 450$	16: $L^* = 500$	15: $L^* = 450$
Минимальное	450	400	400	400	400
Максимальное	500	500	500	500	500
Среднее	481	466	488	462	439
Мода	500	450	500	450	400
Медиана	500	450	500	450	450

Таблица А.10. Значение параметра L^* для ряда dY на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	13: $L^* = 450$	18: $L^* = 500$	11: $L^* = 450$	13: $L^* = 500$	16: $L^* = 400$
Минимальное	400	400	400	400	350
Максимальное	500	500	500	500	500
Среднее	482	468	454	470	409
Мода	500	500	500	500	350
Медиана	500	500	450	500	375

Таблица А.11. Значение параметра r^* для ряда x на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	20: $r^* = 27$	10: $r^* = 26$	13: $r^* = 25$	30: $r^* = 20$	12: $r^* = 17$
Минимальное	25	20	25	20	14
Максимальное	29	26	29	20	17
Среднее	26	23	27	20	15
Мода	26	26	28	20	17
Медиана	26	22	28	20	16

Таблица А.12. Значение параметра r^* для ряда y на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	12: $r^* = 8$	22: $r^* = 23$	13: $r^* = 20$	30: $r^* = 23$	14: $r^* = 29$
Минимальное	8	23	18	23	23
Максимальное	23	29	20	23	29
Среднее	17	23	18	23	25
Мода	23	23	18	23	24
Медиана	17	23	19	23	24

Таблица А.13. Значение параметра r^* для ряда LOD на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	17: $r^* = 15$	30: $r^* = 10$	13: $r^* = 19$	12: $r^* = 19$	15: $r^* = 23$
Минимальное	11	10	12	17	22
Максимальное	15	10	19	27	27
Среднее	12	10	15	22	25
Мода	11	10	19	27	27
Медиана	11	10	17	23	25.5

Таблица А.14. Значение параметра r^* для ряда dX на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	30: $r^* = 3$	30: $r^* = 3$	30: $r^* = 3$	30: $r^* = 3$	30: $r^* = 3$
Минимальное	3	3	3	3	3
Максимальное	3	3	3	3	3
Среднее	3	3	3	3	3
Мода	3	3	3	3	3
Медиана	3	3	3	3	3

Таблица А.15. Значение параметра r^* для ряда dY на 30 последовательных днях.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	17: $r^* = 5$	17: $r^* = 5$	30: $r^* = 3$	27: $r^* = 3$	30: $r^* = 5$
Минимальное	3	3	3	3	5
Максимальное	5	5	3	5	5
Среднее	4	4	3	4	5
Мода	5	5	3	5	5
Медиана	5	5	3	5	5

Таблица А.16. Значение параметра r^* для ряда x на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	4: $r^* = 18$	7: $r^* = 24$	4: $r^* = 10$	6: $r^* = 20$	8: $r^* = 28$
Минимальное	9	9	10	8	8
Максимальное	30	30	30	30	30
Среднее	23	24	22	19	20
Мода	29	29	25	20	8
Медиана	26	26	24.5	20	20

Таблица А.17. Значение параметра r^* для ряда y на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	8: $r^* = 27$	9: $r^* = 17$	6: $r^* = 25$	5: $r^* = 23$	5: $r^* = 23$
Минимальное	5	16	14	7	6
Максимальное	27	29	30	30	29
Среднее	19	23	22	21	19
Мода	23	24	25	23	24
Медиана	23	24	22	22.5	19

Таблица А.18. Значение параметра r^* для ряда LOD на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	11: $r^* = 16$	14: $r^* = 10$	4: $r^* = 14$	6: $r^* = 27$	6: $r^* = 24$
Минимальное	10	10	11	12	17
Максимальное	29	27	29	29	30
Среднее	15	15	19	22	23
Мода	10	10	19	17	19
Медиана	13	12	19	24	23

Таблица А.19. Значение параметра r^* для ряда dX на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	50: $r^* = 3$	50: $r^* = 3$	50: $r^* = 3$	40: $r^* = 3$	10: $r^* = 3$
Минимальное	3	3	3	3	3
Максимальное	3	3	3	5	5
Среднее	3	3	3	3	3
Мода	3	3	3	3	3
Медиана	3	3	3	3	3.5

Таблица А.20. Значение параметра r^* для ряда dY на 50 днях с шагом в 7 дней.

	2006	2007	2008	2009	2010
Максимальная длина неизменяемой последовательности	12: $r^* = 5$	12: $r^* = 3$	12: $r^* = 3$	13: $r^* = 5$	16: $r^* = 5$
Минимальное	3	3	3	3	3
Максимальное	5	5	5	5	5
Среднее	4	4	3	4	4
Мода	5	5	3	5	5
Медиана	5	4	3	5	5

Приложение Б

Графики прогнозов для различных начальных дат

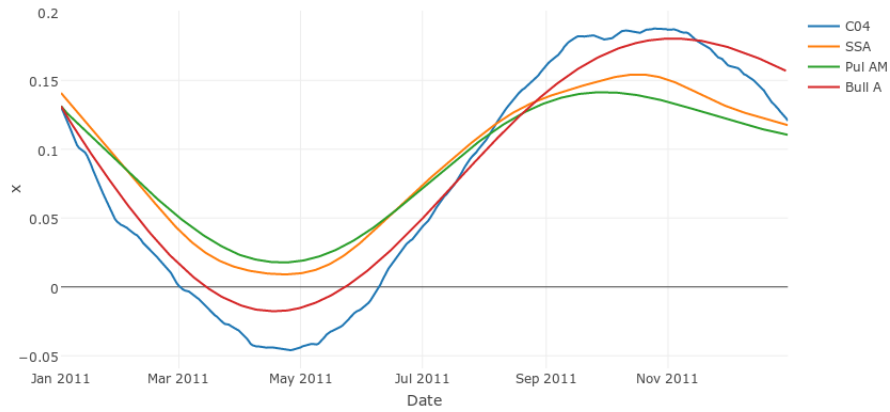


Рис. Б.1. График прогнозов ряда x на 365 дней для MJD = 55562 (1 января 2011 года).

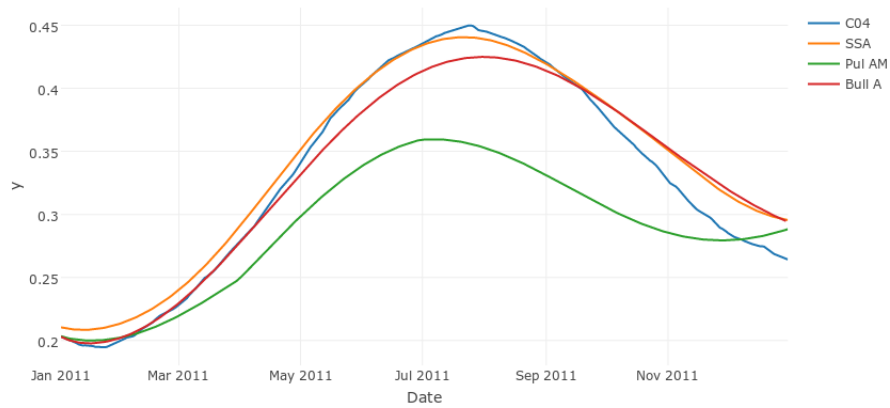


Рис. Б.2. График прогнозов ряда y на 365 дней для MJD = 55562 (1 января 2011 года).

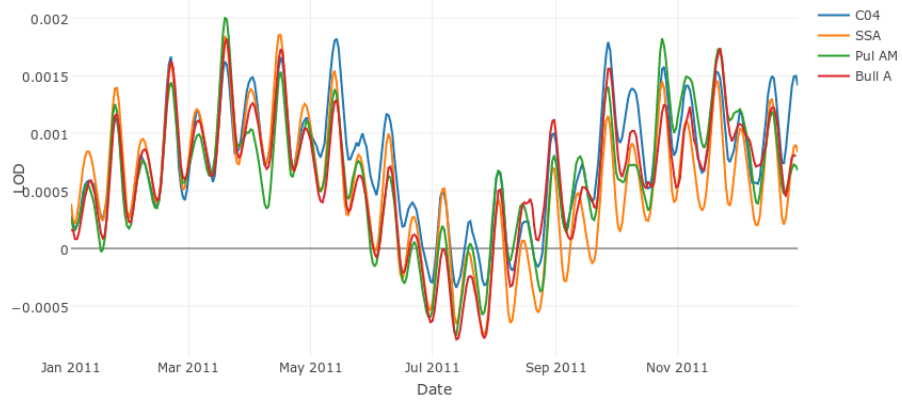


Рис. Б.3. График прогнозов ряда LOD на 365 дней для $MJD = 55562$ (1 января 2011 года).

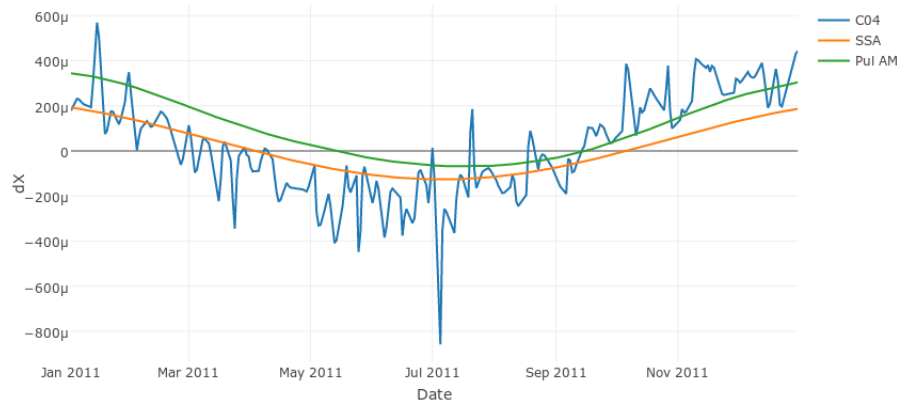


Рис. Б.4. График прогнозов ряда dX на 365 дней для $MJD = 55562$ (1 января 2011 года).

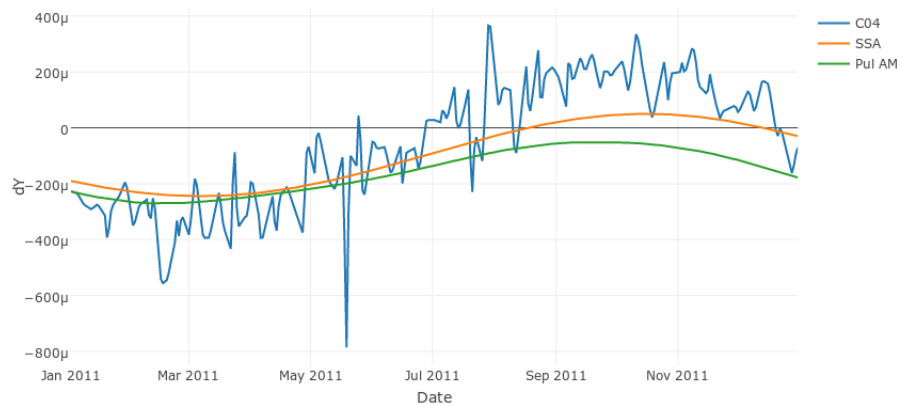


Рис. Б.5. График прогнозов ряда dY на 365 дней для $MJD = 55562$ (1 января 2011 года).

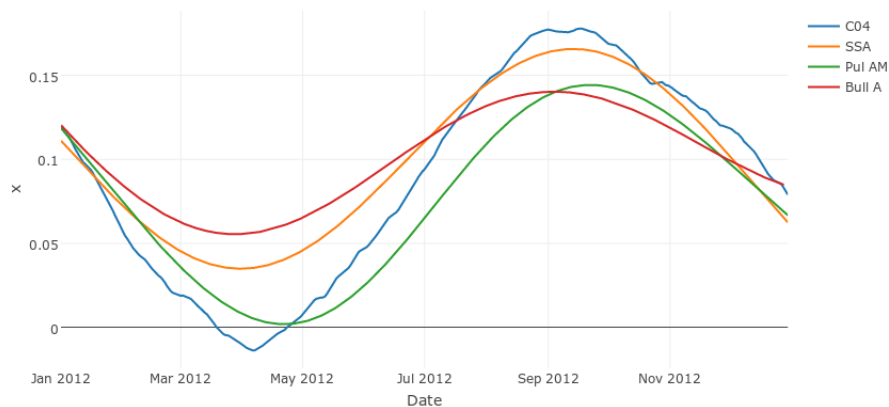


Рис. Б.6. График прогнозов ряда x на 365 дней для MJD = 55927 (1 января 2012 года).

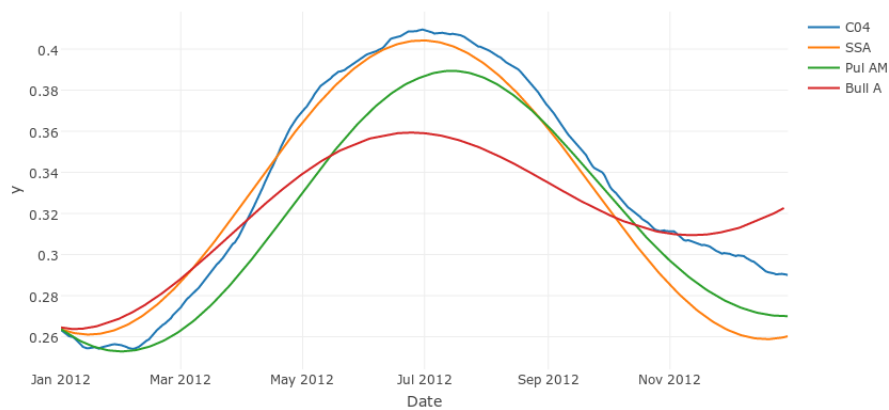


Рис. Б.7. График прогнозов ряда y на 365 дней для MJD = 55927 (1 января 2012 года).

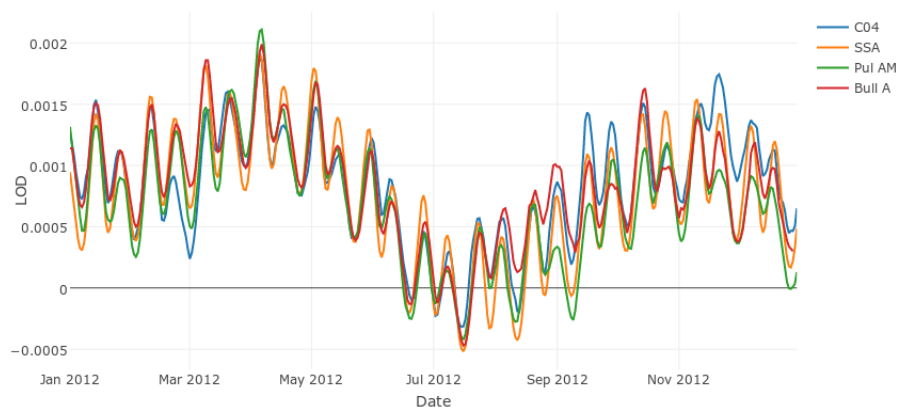


Рис. Б.8. График прогнозов ряда LOD на 365 дней для MJD = 55927 (1 января 2012 года).

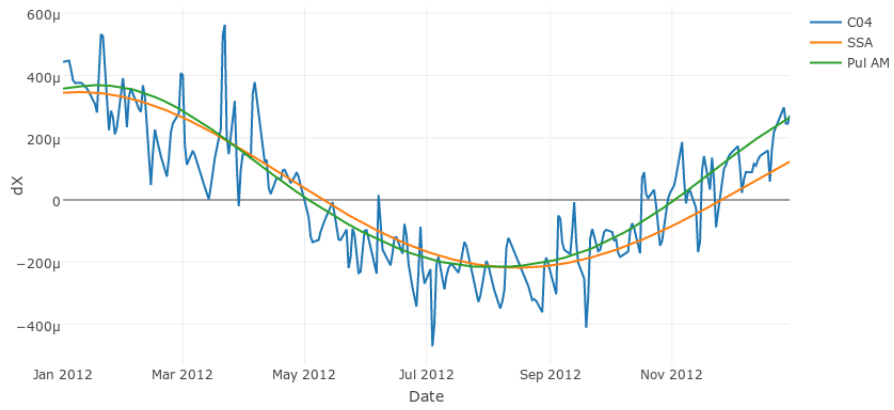


Рис. Б.9. График прогнозов ряда dX на 365 дней для MJD = 55927 (1 января 2012 года).

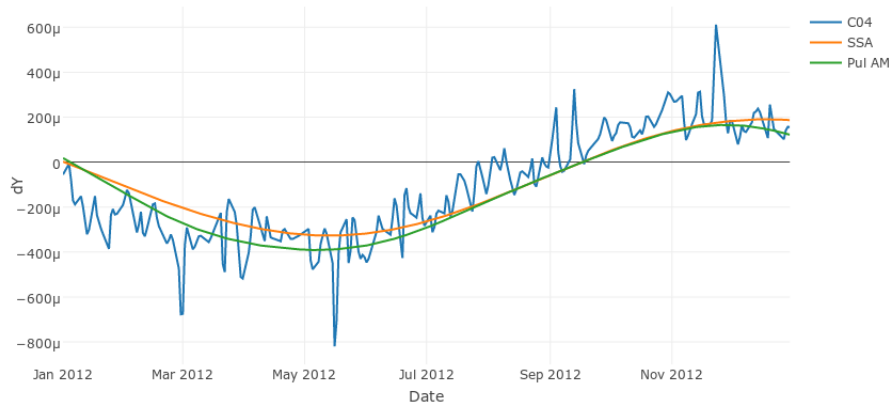


Рис. Б.10. График прогнозов ряда dY на 365 дней для MJD = 55927 (1 января 2012 года).

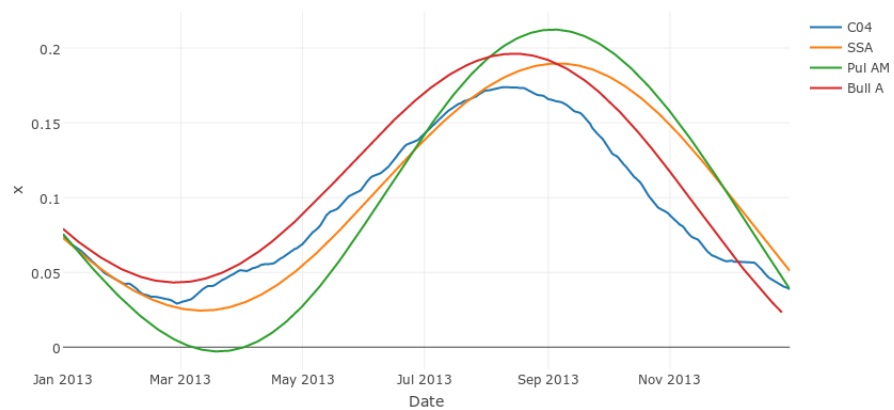


Рис. Б.11. График прогнозов ряда x на 365 дней для MJD = 56293 (1 января 2013 года).

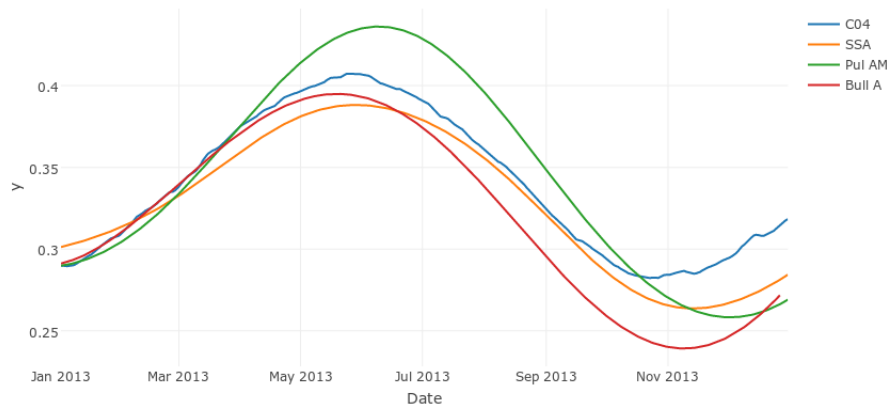


Рис. Б.12. График прогнозов ряда y на 365 дней для $MJD = 56293$ (1 января 2013 года).

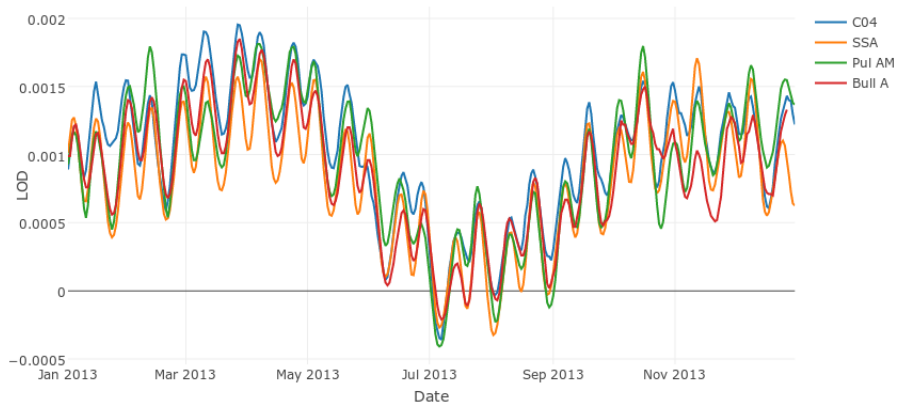


Рис. Б.13. График прогнозов ряда LOD на 365 дней для $MJD = 56293$ (1 января 2013 года).

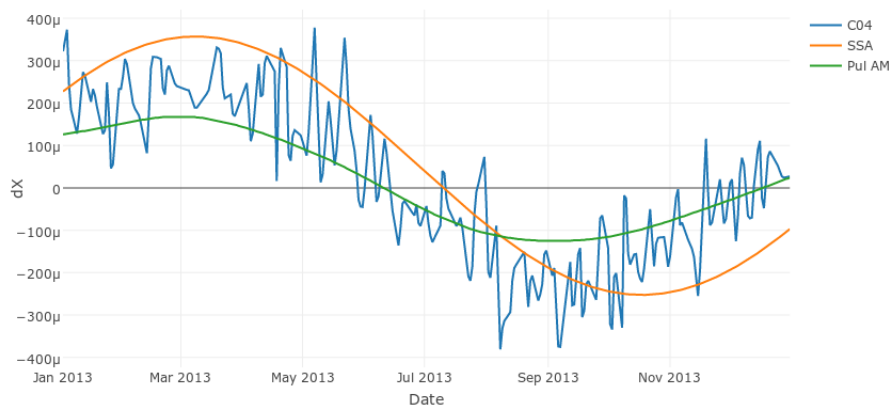


Рис. Б.14. График прогнозов ряда dX на 365 дней для $MJD = 56293$ (1 января 2013 года).

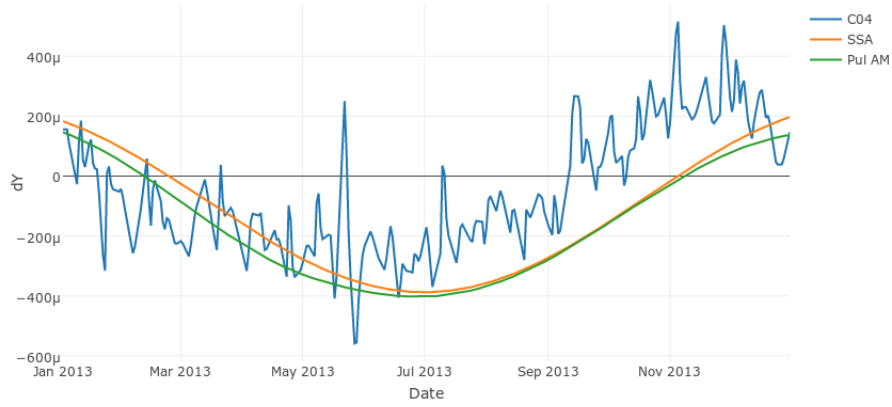


Рис. Б.15. График прогнозов ряда dY на 365 дней для MJD = 56293 (1 января 2013 года).

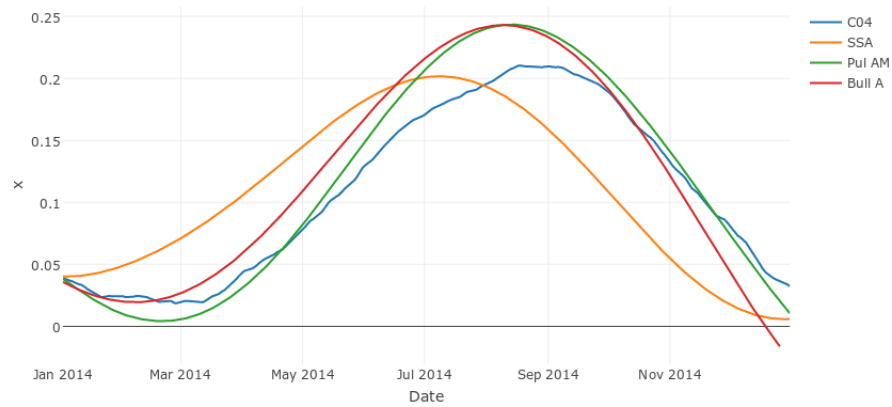


Рис. Б.16. График прогнозов ряда x на 365 дней для MJD = 56658 (1 января 2014 года).

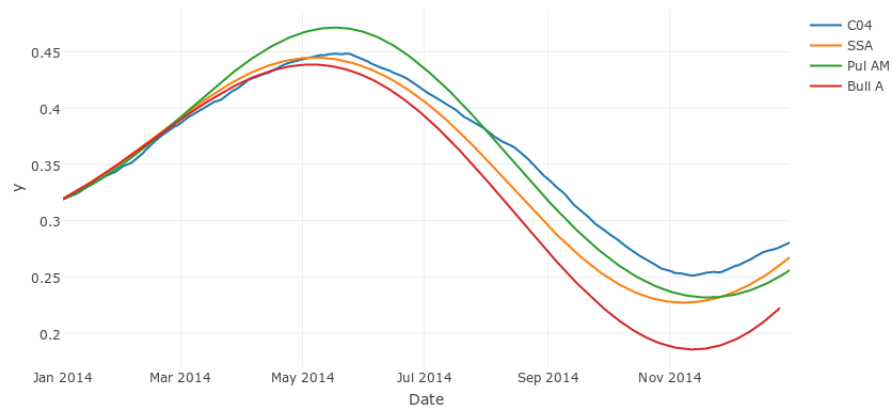


Рис. Б.17. График прогнозов ряда y на 365 дней для MJD = 56658 (1 января 2014 года).

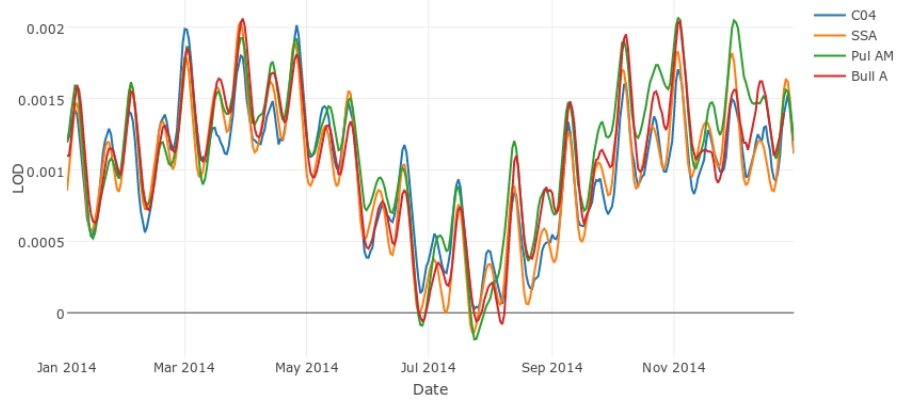


Рис. Б.18. График прогнозов ряда LOD на 365 дней для $MJD = 56658$ (1 января 2014 года).

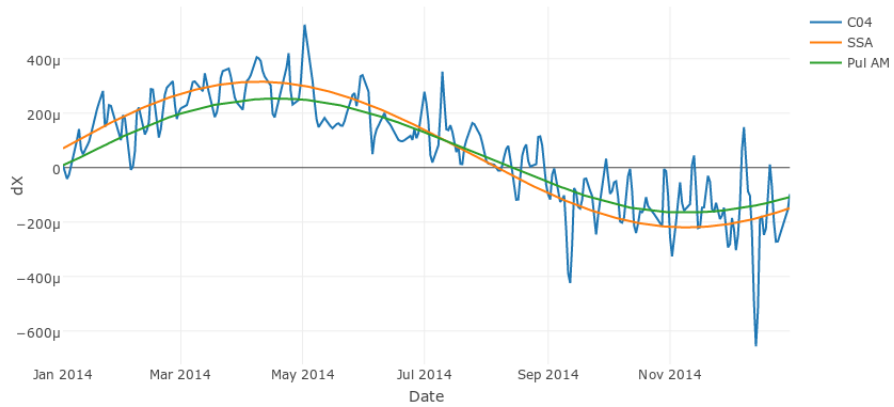


Рис. Б.19. График прогнозов ряда dX на 365 дней для $MJD = 56658$ (1 января 2014 года).

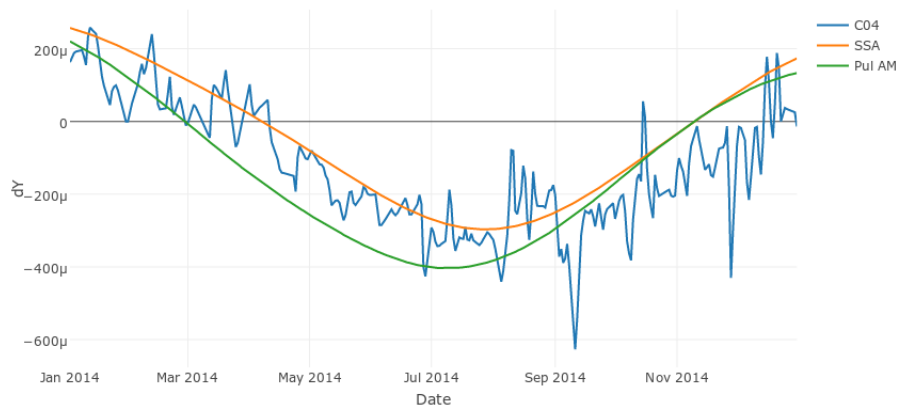


Рис. Б.20. График прогнозов ряда dY на 365 дней для $MJD = 56658$ (1 января 2014 года).